



UNIVERSITAT DE  
BARCELONA

Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques i Informàtica  
Universitat de Barcelona

---

# CAUSALITAT EN INFERÈNCIA ESTADÍSTICA

---

Autor: Pau Satorra Herbera

Director: Dr. Josep Fortiana Gregori  
Realitzat a: Departament de  
Matemàtiques i Informàtica

Barcelona, 18 de gener de 2019

## Abstract

Causal relationships are many times the target of interest in studies on a wide range of fields. In this project, we aim to explain the statistical principles and techniques to infer valid causal effects of one variable over another, in different settings of study designs. We will illustrate these methods by trying to assess the impact of a rise on the minimum wage on employment in fast food restaurants.

## Resum

Les relacions causals són moltes vegades l'objecte d'interès en estudis d'una àmplia varietat de camps. En aquest treball, pretenem descriure els principis i tècniques estadístics per tal d'estimar efectes causals vàlids d'una variable sobre una altra, sota diferents escenaris del disseny de l'estudi. Il·lustrarem aquests mètodes intentant avaluar l'impacte de la pujada del salari mínim en el nombre de treballadors de restaurants de menjar ràpid.

# Agraïments

Agraeixo al meu tutor Josep Fortiana per la guia i els consells que m'ha donat. Al meu pare per haver-me introduït en el món de l'estadística.

# Índex

<b>1</b>	<b>Introducció</b>	<b>1</b>
<b>2</b>	<b>Model de Neyman-Rubin</b>	<b>5</b>
2.1	Unitats múltiples i aleatorització . . . . .	5
2.2	Covariants . . . . .	6
2.3	Mecanisme d'assignació . . . . .	7
2.4	Restriccions en el mecanisme d'assignació . . . . .	8
2.5	Regularitat . . . . .	10
<b>3</b>	<b>Estudis experimentals</b>	<b>12</b>
3.1	Experiments aleatoritzats clàssics . . . . .	12
3.2	p-valors exactes de Fisher . . . . .	13
3.3	Mostreig repetit de Neyman . . . . .	15
3.4	Mètodes de regressió . . . . .	17
3.5	Inferència utilitzant models . . . . .	19
<b>4</b>	<b>Estudis observacionals</b>	<b>22</b>
4.1	Assignació no-confosa . . . . .	22
4.2	Estratificació pel valor de propensió . . . . .	23
4.2.1	Estimació del valor de propensió . . . . .	24
4.2.2	Subclassificació en estrats equilibrats . . . . .	25
4.2.3	Anàlisi de l'efecte causal . . . . .	27
4.3	Emparellament . . . . .	28
4.3.1	Distància de Mahalanobis . . . . .	29
4.3.2	Emparellament i reemplaçament . . . . .	31
4.4	Variància de l'estimador de l'efecte del tractament per emparellament	34
<b>5</b>	<b>Estudi causal sobre l'efecte de la pujada del salari mínim</b>	<b>35</b>
5.1	Assignació no-confosa . . . . .	37
5.2	Estratificació segons el valor de propensió . . . . .	38
5.2.1	Estimació del valor de propensió . . . . .	38
5.2.2	Subclassificació en estrats equilibrats . . . . .	39
5.2.3	Anàlisi causal de l'impacte de la pujada del salari mínim . .	40
5.3	Emparellament . . . . .	41

5.3.1	Distància de Mahalanobis . . . . .	41
5.3.2	Emparellament amb reemplaçament . . . . .	42
<b>6</b>	<b>Conclusions</b>	<b>46</b>
<b>A</b>	<b>Annex</b>	<b>47</b>
A.1	Propietats addicionals . . . . .	47
A.2	Gràfics . . . . .	48

# 1 Introducció

Des de l'origen de la civilització, les relacions de causa-efecte han estat en el centre del pensament humà. Aristòtil, pare de la filosofia i de la ciència, deia en l'Òrganon que tenir el coneixement d'una cosa consistia en saber la causa de la seva existència. En el nostre dia a dia, si observem els problemes amb què ens topem, en la majoria d'ells la pregunta “Per què?” és el primer pas per a la seva solució. Així, camps com la medicina, dret, economia, etc., s'han preocupat en donar una resposta a la pregunta “Per què?” a diferents fenòmens, és a dir en trobar la seva causa. En la pràctica, trobar la causa d'un fenomen és una tasca molt difícil que involucra moltes variables entrelaçades. Així, és més plausible preguntar-se: podem conèixer l'efecte causal d'un fenomen sobre un altre? Per exemple, “Treure l'aspirina el mal de cap?”, “Fumar tabac incrementa el risc a patir càncer de pulmó?”. Donar les eines estadístiques per a respondre a preguntes com aquestes constituirà l'objectiu d'aquest treball.

En general, fins al s.XX, en els llibres d'estadística només s'havia parlat de relacions d'associativitat entre variables, que disten molt de les relacions causals: A i B, dues variables, poden estar associades o correlacionades perquè A causa B, perquè B causa A, o bé perquè una causa externa causa les dues.

Què entenem per efecte causal doncs? Habitualment s'atribueix a Jerzy Neyman (1894-1981) la introducció d'una notació per a l'anàlisi de l'efecte causal, en el context d'un experiment agrícola el 1923. Va plantejar la causalitat des d'una perspectiva contrafactual, punt de vista que ja compartien els científics de l'època. Prenem per exemple l'efecte causal de l'aspirina sobre el mal de cap. Quan diem que l'aspirina ens ha tret el mal de cap, volem dir que si no ens haguéssim pres l'aspirina, encara tindríem mal de cap. En essència doncs, quan afirmem que un fenomen és la causa d'un altre estem dient que si no hagués passat el fenomen que considerem la causa, llavors no hauríem observat el fenomen que considerem l'efecte, això és el que entenem per un punt de vista contrafactual. Neyman va anomenar “potential outcomes” (**resultats potencials**) als dos valors: el què observem si passa la causa, i el què observem si no passa. En el nostre exemple de l'aspirina, els dos resultats potencials serien el mal de cap observat si t'has pres l'aspirina, versus el mal de cap si no te l'has pres. En aquest escenari, s'anomena **tractament** a la variable que conceptualment podem manipular de la qual volem estudiar el seu efecte (l'assignació de l'aspirina). Anomenarem **control** a la unitat particular a la qual no se li ha aplicat el tractament.

Així doncs, per intentar inferir alguna mena d'efecte causal cal conèixer els dos resultats potencials per a poder-los comparar. Si són quantificables, una forma de comparar-los és fent la seva diferència. Com que l'única distinció entre els dos valors dels resultats potencials és l'exposició al tractament, aquesta diferència seria una forma de determinar l'efecte causal del tractament. Notem però, el primer problema que se'ns presenta: ens és impossible observar simultàniament el mal de cap si ens hem pres l'aspirina i si no, ja que no podem, al mateix temps, prendre-la i no prendre-la. Aquest problema es coneix com a **el problema fonamental de la inferència causal**. Si només considerem una unitat, només observarem un

dels dos resultats potencials, de forma que no hi haurà manera de comparar les dues observacions per a determinar-ne l'efecte causal. Ara bé, si tenim més unitats, se'ns obre la possibilitat d'aproximar-lo. Tots els mètodes d'inferència causal acaben sent, més o menys explícitament, mètodes d'imputació dels resultats potencials que falten.

Suposem que tenim més d'una unitat. Podem crear-ne dos grups, a uns els assignem el tractament (grup de tractament) i a uns altres no (grup de control). El procés d'assignació del tractament a les unitats l'anomenarem **mecanisme d'assignació**. Un estimador natural de l'efecte causal, donats aquests dos grups, seria la diferència de les mitjanes dels valors observats entre ells. Ara bé, si els grups construïts, en conjunt tenen atributs diferents, la diferència no solament estimarà l'efecte causal del tractament sinó que també estimarà l'efecte d'altres variables que prenguin valors diferents entre els dos grups. Considerem per exemple dos grups de pacients en els quals a uns se'ls ha administrat l'aspirina i als altres no. Si els pacients escollits per formar part del grup del tractament fossin pacients amb un mal de cap més fort que els del control, fent la diferència no estariem observant l'efecte real de l'aspirina. Necessitem doncs, que els grups que construïm siguin semblants. Anomenarem **covariants** a les variables que representen els atributs de les unitats abans del tractament. Així, per semblants ens referirem al fet que tinguin la mateixa distribució de covariants. Un exemple de covariant en el cas de l'aspirina seria el mal de cap inicial.

La forma d'aconseguir un bon equilibri entre els dos grups és fent l'assignació del tractament aleatòria, com va proposar Neyman el 1936 en el context d'un estudi agrícola. Va plantejar una assignació equivalent a treure boles d'una urna a l'atzar i sense reemplaçament, en la qual hi ha boles vermelles corresponents al tractament i boles blanques corresponents al control. Donada una unitat, si surt una bola vermella se li assigna el tractament, mentres que si surt una bola blanca se li assigna el control. Sota aquest escenari, si agafem la diferència de mitjanes dels valors observats entre els dos grups, aquest és un bon estimador de l'efecte causal. Neyman simplement descriu aquest mecanisme d'assignació (més endavant hi entrarem en més detall) i el posa en pràctica en el context del seu estudi agrícola. Es creu que va ser Ronald Fisher (1890-1962), el 1925, el primer que, independentment de la feina de Neyman va fer un pas més enllà adonant-se de la imprescindibilitat d'aleatoritzar l'assignació per tal de poder fer una inferència causal vàlida.

Neyman i Fisher van proposar solucionar el problema fonamental de la inferència causal a través de l'aleatorització de l'assignació. Hi ha molts estudis però, en els quals no podem aleatoritzar l'assignació, ja sigui per motius logístics o ètics. Per exemple, per intentar inferir l'efecte del tabac en el càncer de pulmó, per raons ètiques no podríem fer fumar a alguns humans i a d'altres no de forma a l'atzar, o bé utilitzar a primats com a objecte d'estudi. Com que seria imprescindible disposar d'animals experimentals amb llarga vida per aquest estudi, només ens quedaria l'opció de fer fumar a les tortugues o als elefants.

Els estudis que sí que podem dissenyar el mecanisme d'assignació s'anomenen **estudis experimentals** mentre que els estudis que no ens és possible controlar el mecanisme d'assignació, o directament ni el coneixem, són el que s'anomenen **es-**

## **tudis observacionals.**

La naturalesa dels estudis observacionals planteja un repte per a la inferència causal. El grup del tractament i el grup de control no són aleatoris, és més, vénen donats, de forma que generalment són grups amb atributs diferents. Comparar-los directament donaria lloc a efectes causals erronis. Per exemple, les farmacèutiques van encarregar fer un estudi per determinar si les dones amb menopausa havien de prendre hormones o no per compensar la seva carència. L'estudi era observacional, es va observar la resistència a malalties de qui prenién les hormones versus qui no en prenien. Comparant els dos resultats, semblava que les hormones tenien un efecte positiu sobre la seva salut. Ara bé, va haver-hi queixes perquè es creia que aquests resultats no reflectien la realitat i l'administració va decidir investigar-ho pel seu compte. Resulta que es va veure que les dones que es prenien les hormones eren precisament les que més es preocupaven per la seva salut en general, les de classe social més alta, les millors alimentades, etc. És a dir les que es prenien les hormones eren les que gaudien d'una millor salut en general, fet que contribuïa a inflar l'efecte positiu de les hormones. De fet, posteriorment es va fer un estudi experimental assignant aleatòriament l'assignació de les hormones, i es va veure que no només no eren beneficioses, sinó que molts cops resultaven ser perjudicials.

No sempre podem reproduir un estudi observacional de forma experimental com en aquest exemple. Com podem inferir efectes causals en el marc d'un estudi observacional doncs? Científics de l'època es limitaven a intentar inferir la causalitat a partir de relacions d'associativitat. Per exemple, R. Doll i A.B. Hill, durant el 1948 i el 1949, es van dedicar a recollir dades de 20 hospitals de Londres per tal d'intentar inferir un efecte causal entre fumar tabac i patir càncer de pulmó. Es tracta d'un estudi observacional, en tant que l'investigador no pot donar de fumar a uns quants individus a l'atzar i a uns altres no, per raons ètiques (no pot dissenyar l'assignació del tractament, fumar tabac), sinó que els fumadors i els no fumadors ja vénen donats. Doll i Hill van deduir de l'estudi una relació d'associativitat i a partir d'aquest resultat van inferir l'efecte causal. Fisher va arremetre en contra de la validesa d'aquest resultat causal, argumentant que l'associativitat no implica la causalitat: podia ser que existís l'efecte causal contrari, és a dir, que el càncer de pulmó provoqués ansietat als pacients de forma que passessin a fumar; o bé, podia haver-hi un factor comú que causés tant el càncer de pulmó com la tendència a fumar, com per exemple una predisposició genètica comuna. Tot i haver-se sabut més tard que Fisher no tenia raó, a més a més que cobrava de les empreses tabaqueres, els seus arguments eren plausibles: era molt difícil inferir la causalitat a partir de l'associativitat de manera correcta, calia una altra forma d'abordar-la en el context d'un estudi observacional.

Donald B. Rubin, el 1974, va proposar un marc per a inferir l'efecte causal, indiferentment de si l'estudi és experimental o observacional, que posava els *potential outcomes* (resultats potencials) en el centre d'aquest, desenvolupant el que es coneix com a model de Neyman-Rubin.

Des d'aleshores, la inferència causal ha esdevingut un camp en continu creixement que ha despertat l'interès d'investigadors de moltes disciplines diferents, des de l'estadística fins a les ciències socials i biomèdiques. Alguns dels autors més prolífics



actualment, a part de Rubin, són: J. Robins, bioestadístic que, entre moltes altres coses, ha desenvolupat mètodes d'inferència causal per a estudis en el que el tractament varia amb el temps; J. Pearl, que aborda la causalitat via models estructurals, perspectiva molt diferent de la de Rubin; S. Greenland, que té una contribució molt important en les limitacions i els mals usos de l'estadística per a anàlisis causals en estudis observacionals; i finalment destacar G.W. Imbens, que aplica la inferència causal, seguint Rubin, en l'econometria. La manera d'enfocar la causalitat i els problemes que aborden varien molt entre els diferents autors. Nosaltres en el treball ens centrarem bàsicament en el punt de vista de Rubin.

En la primera part del treball tractarem estudis experimentals mentre que ja en la segona ens centrarem en els estudis observacionals. Finalment, il·lustrarem les eines estadístiques descrites al llarg del treball reproduint un estudi observacional respecte a l'impacte de la pujada del salari mínim sobre el nombre de treballadors dels restaurants de menjar ràpid de l'estat estatunidenc de Nova Jersey, fet originalment per David Card i Alan B. Krueger el 1993. L'efecte que té la pujada del salari mínim sobre l'economia és un tema del qual es parla molt en l'actualitat, després que el govern Espanyol aprovés la pujada del salari mínim fins als 900 € mensuals, sumat a la concessió que va fer Macron de pujar-lo 100 € per les protestes dels *gilets jaunes*. Són molts els economistes que, basant-se en els seus models econòmics, afirmen que la pujada té un efecte negatiu sobre l'economia, en el sentit que les empreses es veuen forçades a reduir el nombre de treballadors, i per tant, la seva demanda baixa. Alguns d'altres, com el premi Nobel d'Economia Joseph Stiglitz, s'han pronunciat en la direcció contrària, recolzant-se en diversos estudis estadístics en els quals no es troba un efecte negatiu significatiu o fins i tot es troba un de positiu. Aquest últim resultat va ser el que van trobar Card i Krueger en el seu estudi, utilitzant mètodes que no tenien en compte la naturalesa observacional de l'estudi. A quines conclusions arribarem nosaltres utilitzant els mètodes que descriurem al llarg del treball en el context d'un estudi observacional?

Comencem el treball introduint el que es coneix com a model de Neyman-Rubin que més que un model, és un marc de referència per abordar el mecanisme d'assignació.

## 2 Model de Neyman-Rubin

En tot aquest capítol suposarem que volem estudiar l'efecte causal d'un tractament sobre una variable  $Y$  d'una població finita de  $N$  individus.

### 2.1 Unitats múltiples i aleatorització

Com ja hem explicat, per a poder inferir alguna mena d'efecte causal, que és la comparació entre els dos resultats potencials, una sola observació en una sola unitat és insuficient. Una opció podria ser observar la mateixa unitat en dos temps diferents: en un li donem el tractament i en l'altre el control. Però el temps és important de forma que l'efecte causal que puguem observar pot ser degut al pas del temps entre les dues observacions, i no al tractament en si. Així doncs, la millor opció serà observar moltes unitats al mateix temps, unes exposades al tractament i altres no. És necessari fer algunes suposicions abans per a preservar la independència de les unitats.

Cal, en primera instància, que els resultats potencials de cada unitat no variïn amb el tractament assignat a altres unitats. Considerant l'exemple de l'aspirina, si tanquem en la mateixa sala d'espera un pacient tractat i un pacient no tractat, pot ser que aquest últim, a causa del mal de cap, es queixi contínuament i molesti el primer passant a tenir mal de cap ell també. Així doncs, el mal de cap del primer pacient no seria degut al seu propi tractament, sinó al tractament del pacient del costat, fet que portaria a un error en l'anàlisi causal. En segona instància, és important evitar diferents versions del tractament per a diferents unitats. Estaríem aleshores estudiant efectes causals sota diferents tractaments, que no és el que ens hem plantejat. Aquestes dues suposicions formen el que s'anomena "The Stable Unit Treatment Value Assumption" (**SUTVA**) i són indispensables per a la validesa de la inferència causal. D'ara en endavant, suposarem sempre que SUTVA és preservat.

Reprenem l'exemple de l'aspirina i el mal de cap amb més profunditat per a il·lustrar el paper de l'aleatorització en la inferència causal quan tenim més d'una unitat. Suposem que tenim una sèrie d'observacions sobre diferents pacients amb mal de cap en els quals a alguns se'ls ha donat una aspirina i a d'altres no. Denotem per  $Y(0)$  i  $Y(1)$  els vectors que tenen per components el mal de cap de cada pacient sota el control i sota l'efecte de l'aspirina respectivament. Per al pacient  $i$ -èssim, només hem observat  $Y_i(1)$  o bé  $Y_i(0)$ . Per exemple, tindríem una taula de la següent forma:

$Y(0)$	$Y(1)$	$Y(0)$ vs $Y(1)$
$Y_1(0)$	?	?
?	$Y_2(1)$	?
?	$Y_3(1)$	?
$Y_4(0)$	?	?
$\vdots$	$\vdots$	$\vdots$

Volem comparar l'efecte de l'aspirina en el conjunt dels pacients:  $Y(0)$  versus  $Y(1)$ . Per exemple, el podríem determinar per la diferència de mitjanes  $\bar{Y}(1) - \bar{Y}(0)$ . Però només tenim una observació per a cada pacient  $i$ -èssim:  $Y_i(1)$  o bé,  $Y_i(0)$ , el que constitueix el problema fonamental de la inferència causal. Així, no coneixem ni  $\bar{Y}(1)$  ni  $\bar{Y}(0)$ . Si tots els pacients fossin molt semblants, podríem comparar l'observació sota el tractament d'un pacient amb l'observació sota el control d'un altre obtenint l'efecte de l'aspirina, ja que totes les observacions les podríem tractar com si vinguessin del mateix pacient. D'aquesta forma, la diferència de les mitjanes entre els valors observats sota el tractament i sota el control  $\bar{Y}^{\text{obs}}(1) - \bar{Y}^{\text{obs}}(0)$  seria un bon estimador de l'efecte causal  $\bar{Y}(1) - \bar{Y}(0)$ .

Aquesta situació és ideal, en la pràctica, dos pacients tindran unes característiques diferents que els faran reaccionar de forma diferent quan es prenen l'aspirina, és a dir, les unitats són heterogènies. Tot i això, fent algunes suposicions extremes,  $\bar{Y}^{\text{obs}}(1) - \bar{Y}^{\text{obs}}(0)$  pot seguir sent un bon estimador de l'efecte causal. El cas més simple és suposant que hem dissenyat l'experiment de forma que l'assignació de l'aspirina és completament a l'atzar. En aquest escenari, el grup del tractament i el grup del control hauran estat seleccionats aleatòriament, de forma que si ens els mirem en conjunt, seran dos grups semblants, amb característiques distribuïdes de forma similar. Així doncs,  $\bar{Y}^{\text{obs}}(1) - \bar{Y}^{\text{obs}}(0)$  ens aproximarà l'efecte causal de l'aspirina, ja que en global els dos grups només es diferencien en l'assignació de l'aspirina. Una altra forma de pensar el paper de l'aleatorització és mirar-nos l'experiment com si fos un mostreig aleatori, és a dir, imaginant-nos que el grup del tractament i del control són dues mostres d'una població de pacients. És sabut en teoria de mostreig que si les mostres són triades a l'atzar i són prou grans representaran a tota la població. Així doncs, podem prescindir dels valors no observats tant en  $Y(1)$  com en  $Y(0)$ , restringint-nos únicament als resultats observats per a extrapolar conclusions sobre l'efecte causal de l'aspirina en el conjunt dels pacients.  $\bar{Y}^{\text{obs}}(1)$  ens aproximarà  $\bar{Y}(1)$  i  $\bar{Y}^{\text{obs}}(0)$  ens aproximarà  $\bar{Y}(0)$ , de manera que  $\bar{Y}^{\text{obs}}(1) - \bar{Y}^{\text{obs}}(0)$  serà un bon estimador de  $\bar{Y}(1) - \bar{Y}(0)$ .

Més endavant, un cop formalitzades les nocions bàsiques, entrarem amb més detall en aquesta forma d'inferir la causalitat, que es coneix com a mostreig repetit de Neyman, i en moltes altres més en aquest context d'un estudi experimental amb l'assignació aleatoritzada.

## 2.2 Covariants

La majoria de les vegades la variable  $Y$  depèn de molts altres factors a part del tractament del qual volem estudiar-ne l'efecte causal. Pot dependre de molts altres atributs dels individus que, si els coneixem, ens ajudaran a caracteritzar-los molt més. Aquestes variables les anomenarem covariants o variables prèvies al tractament. Si tenim  $K$  covariants, denotarem per  $X$  a la matriu de dimensió  $N \times K$  que recull els valors que prenen aquestes covariants en tots els individus de la població.

Les covariants poden facilitar la inferència causal de tres formes diferents. En primer

lloc, poden servir per fer estimacions més precises sobre l'efecte causal. Per exemple, en el cas de l'aspirina, si tenim en compte el mal de cap abans del tractament a l'hora de determinar l'efecte de l'aspirina (podríem calcular l'efecte fent la diferència entre el mal de cap final i el inicial) aïllaré millor aquest efecte. En segon lloc, podem estar interessats en l'efecte causal de subgrups de la població, definits segons unes covariants. Per exemple, si volguéssim estudiar l'efecte de l'aspirina per homes i per dones separatament. Finalment, l'ús més important de les covariants és el d'evitar la dependència del mecanisme d'assignació respecte dels resultats potencials. Més endavant descriurem aquest últim ús.

Per a poder usar les covariants adequadament, és important que el tractament no les modifiqui, ja que són observades prèviament al tractament. També és important que durant l'espai temporal entre l'observació de les covariants i l'observació dels resultats potencials, aquestes no canviïn el seu valor.

En estudis socials, les covariants solen incloure el sexe, l'edat, l'educació, l'estatus econòmic, etc. En estudis biomèdics, sovint inclouen la història mèdica de l'individu, els antecedents familiars, etc.

## 2.3 Mecanisme d'assignació

Introduïm la notació que va plantejar Rubin per abordar el problema, que generalitza el que teníem en l'exemple de l'aspirina per tal que també serveixi per estudis observacionals. Considerem la variable  $Y$  sobre la qual volem determinar-ne l'efecte causal d'un tractament. Com hem fet en l'exemple de l'aspirina, anomenarem  $Y(0)$  i  $Y(1)$  als dos vectors corresponents als resultats potencials de les  $N$  unitats, sota el control i sota el tractament respectivament. És a dir,  $Y_i(0)$  serà el valor de la variable  $Y$  de la unitat  $i$ -èssima sota el control. Igual per  $Y_i(1)$  però sota el tractament. Denotem per  $W$  al vector d'assignacions, que serà un vector de  $N$  components que adoptarà zeros o uns en funció de què se li ha assignat a la unitat corresponent. Per exemple, si a la unitat  $i$ -èssima se li ha assignat el tractament tindrem  $W_i = 1$ , mentre que si se li aplica el control  $W_i = 0$ . Per tant,  $W$  adoptarà valors sobre  $\{0, 1\}^N$ . Així, podem definir els vectors  $Y^{\text{obs}}$  i  $Y^{\text{fal}}$  que per a cada compomenen tenen respectivament el resultat observat:

$$Y_i^{\text{obs}} = Y_i(W_i) = \begin{cases} Y_i(0) & \text{si } W_i = 0, \\ Y_i(1) & \text{si } W_i = 1 \end{cases}$$

i el resultat potencial no observat que falta:

$$Y_i^{\text{fal}} = Y_i(1 - W_i) = \begin{cases} Y_i(1) & \text{si } W_i = 0, \\ Y_i(0) & \text{si } W_i = 1 \end{cases}$$

Considerem la matriu  $X$  de les covariants, de dimensió  $N \times K$ . Els  $N$  vectors de les seves files que denotarem per  $X_i$  correspondran a les  $K$  covariants que caracteritzen a la unitat  $i$ -èssima. Donats els valors de  $X, Y(0)$  i  $Y(1)$ , el mecanisme d'assignació és la funció que assigna una probabilitat a tots els  $2^N$  valors possibles que pugui adoptar el vector d'assignacions  $W$ :

**Definició 2.1.** *El mecanisme d'assignació és una distribució de probabilitat  $P$  sobre l'espai  $\Omega = \{0, 1\}^N$ , condicionada als valors de  $X$ ,  $Y(0)$  i  $Y(1)$ :*

$$P: \Omega \rightarrow [0, 1]$$

$$W \mapsto P(W | X, Y(0), Y(1))$$

El fet que en principi l'assignació del tractament a cada unitat pugui dependre dels resultats potencials, a priori desconeguts, és el que ens permet generalitzar el mecanisme d'assignació pels estudis observacionals. Més endavant veurem que la idea clau per a la validesa de la inferència causal serà precisament que el mecanisme d'assignació no ens acabi dependent d'aquests resultats potencials.

El mecanisme d'assignació està definit de forma conjunta per a totes les unitats. En particular, la probabilitat de cada unitat de ser assignada al tractament és:

**Definició 2.2.** *La probabilitat d'assignació d'una unitat  $i$ -èssima és la suma de les probabilitats de les combinacions de  $W$  que assignin el tractament a la unitat  $i$ -èssima, és a dir:*

$$p_i(X, Y(0), Y(1)) = \sum_{W: W_i=1} P(W | X, Y(0), Y(1))$$

La probabilitat de ser assignada al tractament de control serà doncs:

$$1 - p_i(X, Y(0), Y(1))$$

Sovint, ens serà útil el promig de les probabilitats d'assignació de totes les unitats que tinguin en comú un mateix valor de les covariants  $X_i = x$ . Així definim:

**Definició 2.3.** *El **valor de propensió** (“propensity score”) en  $x$  és el promig de les probabilitats d'assignació de les unitats amb un mateix valor de les covariants  $X_i = x$ :*

$$e(x) = \frac{1}{N(x)} \sum_{i: X_i=x} p_i(X, Y(0), Y(1))$$

on  $N(x) = \#\{i \in \{1, \dots, N\} | X_i = x\}$ . Per valors de  $x$  tals que  $N(x) = 0$  definirem  $e(x) = 0$ .

## 2.4 Restriccions en el mecanisme d'assignació

Hi ha una sèrie de propietats que generalment demanarem que compleixi el mecanisme d'assignació per tal de fer possible la inferència causal.

La primera propietat és que l'assignació d'una unitat no depengui dels resultats potencials o de l'assignació de les altres unitats. Si el mecanisme d'assignació satisfà aquesta propietat, s'anomenarà individualístic:

**Definició 2.4.** Un mecanisme d'assignació  $P(W|X, Y(0), Y(1))$  és individualístic si existeix alguna funció  $q(\cdot) \in [0, 1]$  tal que la probabilitat d'assignació de cada unitat satisfaci:

$$p_i(X, Y(0), Y(1)) = q(X_i, Y_i(0), Y_i(1))$$

per a tot  $i \in \{1, \dots, N\}$  i també compleixi:

$$P(W|X, Y(0), Y(1)) = c \prod_{i=1}^N q(X_i, Y_i(0), Y_i(1))^{W_i} (1 - q(X_i, Y_i(0), Y_i(1)))^{1-W_i}$$

per  $(W, X, Y(0), Y(1)) \in \mathbb{A}$ , per algun conjunt  $\mathbb{A}$ , i és zero a la resta.  $c$  és una constant, assegura que la suma de totes les probabilitats sigui 1.

La segona propietat és que potencialment cada unitat pugui ser assignada al grup de tractament o al grup de control. És a dir no pot haver-hi cap  $i$  tal que  $p_i(X, Y(0), Y(1)) \in \{0, \}$ . Parlarem aleshores d'assignació probabilística:

**Definició 2.5.** Un mecanisme d'assignació és probabilístic si la probabilitat d'assignació de cada unitat està estrictament entre 0 i 1:

$$0 < p_i(X, Y(0), Y(1)) < 1$$

per a cada  $\{X, Y(0), Y(1)\}$  possible.

Recordem que de la manera en la qual hem definit el mecanisme d'assignació, la probabilitat d'assignació del tractament podria dependre dels resultats potencials. Quan això passa direm que el mecanisme d'assignació és confós. La tercera i última propietat d'un mecanisme d'assignació consisteix doncs a restringir aquesta dependència únicament a les covariants. Tindrem aleshores un mecanisme d'assignació no-confós (“unconfounded assignment”):

**Definició 2.6.** Un mecanisme d'assignació és **no-confós** (“unconfounded”) si:

$$P(W|X, Y(0), Y(1)) = P(W|X)$$

per a tot  $W, X, Y(0), Y(1)$ .

Equivalentment, si ens ho mirem unitat a unitat, utilitzant la notació de Dawid (1979) per a la independència condicionada podem reescriure aquesta propietat com:

$$W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid X_i$$

per a totes les unitats  $i \in \{1, \dots, N\}$ .

D'alguna manera, amb la propietat de no-confusió estem demanant que les unitats tractades no difereixin de les controls en característiques que no puguem controlar a través d'unes covariants i afectin els resultats potencials.

Gràcies a l'aleatorització del mecanisme d'assignació, si tenim suficients unitats en cada grup, aconseguim que la distribució dels atributs siguin iguals i al seu

torn les unitats semblants entre els dos grups, com ja hem explicat. Per tant, tindrem automàticament que el mecanisme d'assignació és no-confós. De fet tenim inclús que  $P(W|X) = P(W)$ , ja que l'assignació és completament aleatòria, no dependrà de cap covariant. Però aleatoritzar només ho podem fer en el marc d'un estudi experimental, quan nosaltres controlem el mecanisme d'assignació. En un estudi observacional, no tindrem unitats tractades i no tractades a l'atzar. Per exemple, les persones que fumen són normalment persones que també beuen, i que tenen hàbits poc saludables en general. Per tant, el grup d'individus en el tractament (els fumadors) tindran característiques totalment diferents del grup del control (no fumadors), en el context d'un estudi observacional de l'efecte del tabac sobre la mortalitat. Aquestes característiques, al seu torn, estan associades a la mortalitat i per tant, a priori, generalment en un estudi observacional tenim un mecanisme d'assignació que depèn dels resultats potencials. Però si som capaços de controlar totes aquestes característiques incloent-les com a covariants tindrem un mecanisme d'assignació que ja no dependrà dels resultats potencials sinó de les covariants, aconseguint que el mecanisme d'assignació sigui no-confós. Controlar aquestes diferències serà essencial per a fer una inferència causal vàlida en el context d'un estudi observacional, com veurem més endavant. La clau és doncs que les variables que prenguin valors diferents entre els dos grups entrin com a covariants. Aquest és l'ús més important de les covariants.

Observem que el valor de propensió d'un mecanisme d'assignació no-confós i individualístic es correspon amb la probabilitat d'assignació de qualsevol unitat  $i$ -èssima tal que  $X_i = x$ :

$$e(x) = \frac{1}{N_x} \sum_{i: X_i=x} q(X_i) = q(x)$$

D'acord amb aquestes restriccions tindrem dos tipus de mecanisme d'assignació:

## 2.5 Regularitat

Tant en estudis observacionals com en experimentals, definim el mecanisme d'assignació regular:

**Definició 2.7.** *Un mecanisme d'assignació és **regular** si:*

1. *És individualístic*
2. *És probabilístic*
3. *És no-confós*

Si el mecanisme d'assignació no és regular, direm que és irregular. Al llarg del treball ens centrarem amb mecanismes d'assignació regulars ja que la irregularitat complica molt la validesa de la inferència causal. A més a més, normalment en estudis observacionals ben dissenyats, els investigadors han recollit totes les covariants importants associades tant amb els resultats potencials com amb l'assignació del

tractament, de manera que no és desbaratat aproximar el mecanisme d'assignació real per un de regular.

Comencem abordant els estudis experimentals que tenen un mecanisme d'assignació regular, que Neyman ja va descriure el 1923.



### 3 Estudis experimentals

Mantenim la mateixa notació que hem introduït en l'anterior capítol.

#### 3.1 Experiments aleatoritzats clàssics

En estudis experimentals, en els quals controlem el mecanisme d'assignació, podem dissenyar l'assignació de forma que s'apliqui el tractament o el control de forma aleatòria d'entre la població. És a dir, seleccionem  $W$  a l'atzar d'entre el conjunt de tots els possibles valors  $\Omega = \{0, 1\}^N$ . D'aquesta manera, com ja hem explicat abans, tindrem directament que el mecanisme d'assignació és no-confós. És clar que serà també individualístic i probabilístic, de forma que tindrem que és regular. Els estudis experimentals amb el mecanisme d'assignació regular els anomenarem experiments aleatoritzats clàssics.

Podem tenir diferents tipus d'experiments aleatoritzats clàssics basant-nos en la restricció dels possibles valors que pot adoptar el mecanisme d'assignació  $W$ . Sigui  $\Omega^+ \in \Omega = \{0, 1\}^N$  el subconjunt de valors de  $W$  sobre els quals restringim el nostre experiment, o equivalentment, amb probabilitat positiva. Fixem-nos que hi ha combinacions de  $W$  que no ens donen cap mena d'informació sobre l'efecte causal que volem inferir. Per exemple, assignacions en el que totes les unitats reben el tractament impossibilita la comparació dels resultats potencials. Tenint en compte això, ens centrarem en els experiments completament aleatoritzats:

**Definició 3.1.** *Un **experiment completament aleatoritzat** és un experiment aleatoritzat clàssic tal que es fixa a priori el nombre  $N_t \in \{1, \dots, N-1\}$  d'unitats que seran assignades al tractament. Les  $N_t$  unitats seran triades a l'atzar d'entre tota la població.*

Notem que d'aquesta forma evitem precissament combinacions com per exemple la que correspon a assignar a totes les unitats el tractament. De fet, el subconjunt  $\Omega^+$  d'assignacions amb probabilitat positiva serà:

$$\Omega^+ = \{W \in \Omega \mid \sum_{i=1}^N W_i = N_t\}$$

on  $N_t \in \{1, 2, \dots, N-1\}$ .

Bàsicament, qualsevol experiment completament aleatoritzat és equivalent al de treure boles d'una urna sense reemplaçament amb  $N_t$  boles d'un color que correspondran al tractament i  $N_c$  boles d'un altre que correspondran al control. Cada bola que traiem correspondrà a l'assignació que farem a una unitat. És a dir, les  $N_t$  unitats a les quals els hi seran assignades el tractament s'escullen a l'atzar d'entre totes les  $N$  unitats de la població. Per tant, la cardinalitat del subconjunt  $\Omega^+$  és el nombre combinatori  $\binom{N}{N_t}$ . Tots els valors  $W \in \Omega^+$  són igualment probables de forma que la probabilitat de cada un és  $\binom{N}{N_t}^{-1}$ .

Així doncs, el mecanisme d'assignació d'un experiment completament aleatoritzat tindrà la següent forma:

$$P(W|X, Y(0), Y(1)) = P(W) = \begin{cases} \left(\frac{N}{N_t}\right)^{-1}, & \text{si } W \in \Omega^+ \\ 0, & \text{altrament} \end{cases}$$

La primera igualtat la tenim perquè el mecanisme d'assignació és totalment a l'atzar, no depèn ni de les covariants.

Altres formes d'experiments aleatoritzats clàssics sortirien de restringir encara més el subconjunt  $\Omega^+$ . Un experiment aleatoritzat estratificat consisteix en dividir la població en estrats segons covariants semblants i en cada un d'ells realitzar un experiment completament aleatoritzat. Finalment, si imposem que els estrats tinguin dues unitats tindrem el que s'anomena un experiment aleatoritzat emparellat. Com que al cap i a la fi s'acaben reduint a fer un experiment completament aleatoritzat, els mètodes d'inferència causal són molt semblants, de forma que ens centrarem únicament en la inferència causal en el context d'un experiment completament aleatoritzat.

En les següents seccions descriurem les diverses formes d'inferir l'efecte causal d'un tractament sobre una variable  $Y$ , donat un experiment completament aleatoritzat:

## 3.2 p-valors exactes de Fisher

Fisher va enfocar el problema plantejant una hipòtesi nul·la sobre l'efecte causal: no hi ha efecte entre el tractament i el control. És a dir, sota la hipòtesi nul·la, totes les unitats tenen els dos resultats potencials idèntics. No tenim per què creure en la hipòtesi nul·la, de fet es tracta de veure amb quina significació les mateixes dades parlen en contra, vegem-ne com:

Considerem qualsevol estadístic de test  $T$ , funció del vector estocàstic d'assignació  $W$ , de  $Y^{\text{obs}}$  i de les covariants  $X$ :  $T(W, Y^{\text{obs}}, X)$ . Notem que la tria de  $W$  és aleatòria, de forma que  $Y^{\text{obs}}$  també ho serà i al seu torn  $T$  també. Considerem el  $T^{\text{obs}}$  resultant del nostre experiment completament aleatoritzat donat el  $W$  seleccionat a l'atzar. Considerem la hipòtesi nul·la que ens diu que el tractament no té cap efecte:

$$H_0 : Y_i(0) = Y_i(1)$$

Sota aquesta hipòtesi, podem calcular tots els valors de  $T$  per a cada vector d'assignació  $W$  aleatori possible perquè per a cada unitat podem imputar el resultat potencial que falta, ja que tenim que  $Y_i(0) = Y_i(1) = Y_i^{\text{obs}}$ . Calculada la distribució de l'estadístic sota la hipòtesi nul·la ens podem preguntar: quant inusual és el  $T^{\text{obs}}$  realment observat assumint que la hipòtesi nul·la és verdadera? Una forma de veure-ho és preguntant-nos quant de probable és observar un valor de l'estadístic que sigui tan gran com l'observat realment  $T^{\text{obs}}$ , segons la distribució calculada. Això ens ho dirà el p-valor, que és  $Pr(T \leq T^{\text{obs}} | H_0)$  o  $Pr(T \geq T^{\text{obs}} | H_0)$  segons el valor  $T^{\text{obs}}$  sigui extrem per l'esquerra o per la dreta respectivament. En teoria, si efectivament se satisfés la hipòtesi nul·la, el valor de  $T^{\text{obs}}$  no hauria de ser massa

extrem d'entre tots els valors diferents de  $T$  que vénen donats per diferents  $W$ , ja que el  $W$  del nostre experiment ha estat escollit a l'atzar.

Per exemple, podem agafar la diferència dels promitjos dels valors observats pel tractament i pel control:

$$T^{dif} = \left| \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \right|$$

on:

$$\overline{Y}_c^{\text{obs}} = \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{\text{obs}} \quad \text{i} \quad \overline{Y}_t^{\text{obs}} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{\text{obs}}$$

És important que l'estadístic test sigui sensible en vers a allunyaments respecte de la hipòtesi nul·la. Així doncs, per a determinar quin estadístic utilitzem cal considerar totes les hipòtesis alternatives. Per exemple, si creiem que l'efecte causal pot ser multiplicatiu, el següent estadístic adoptarà valors més grans en el cas que sigui certa la hipòtesi alternativa:

$$T^{\log} = \left| \frac{\sum_{i:W_i=1} \ln(Y_i^{\text{obs}})}{N_t} - \frac{\sum_{i:W_i=0} \ln(Y_i^{\text{obs}})}{N_c} \right|$$

Hi ha una classe important d'estadístics que no són funció del valor del resultat, sinó del seu rang dins d'una ordenació de més petit a més alt. Són els estadístics rangs. Definim el rang de la unitat  $i$ -èssima, en l'absència de valors empatats com la posició que ocupa el seu resultat potencial observat:

$$\tilde{R}_i = \sum_{j=1}^N \mathbb{1}_{Y_j^{\text{obs}} \leq Y_i^{\text{obs}}}$$

Per obtenir un rang normalitzat que tingui mitjana zero en la mostra cal restar-li  $(N+1)/2$ .

Si hi haguessin valors empatats, llavors aquesta definició no ens val. Suposem que dos valors fossin idèntics a  $y$  i que sota d'aquest valor  $y$  hi hagués  $L$  unitats. Una opció seria assignar a una el rang  $L+1$  i a l'altra  $L+2$ . El que farem serà donar el rang promig a les dues:  $((L+1) + (L+2))/2$ . Per tant, el rang normalitzat de la unitat  $i$ -èssima d'una mostra amb valors empatats és:

$$R_i = \sum_{j=1}^N \mathbb{1}_{Y_j^{\text{obs}} < Y_i^{\text{obs}}} + \frac{\sum_{j=1}^N \mathbb{1}_{Y_j^{\text{obs}} = Y_i^{\text{obs}}}}{2} - \frac{N+1}{2}$$

Tots aquests estadístics proposats ignoren la presència de les covariants. Si les volguéssim tenir en compte per tal d'aïllar millor l'efecte que és purament del tractament, hauríem d'utilitzar un estadístic que fos funció d'elles. Per exemple, podem agafar la diferència de les mitjanes dels guany:

$$T(W, Y_{\text{obs}}, X) = T^{\text{guany}} = |(\overline{Y}_t^{\text{obs}} - \overline{X}_t) - (\overline{Y}_c^{\text{obs}} - \overline{X}_c)|$$

on  $\overline{Y_t^{\text{obs}}} = \sum_{i:W_i=1} Y_i^{\text{obs}}/N_t$ ;  $\overline{Y_c^{\text{obs}}}$  igual per als controls;  $\overline{X_t} = \sum_{i:W_i=1} X_i/N_t$ ; i  $\overline{X_c}$  igual per als controls.

### 3.3 Mostreig repetit de Neyman

El plantejament de Neyman, com hem introduït en l'exemple de l'aspirina, consisteix a veure que la diferència de mitjanes dels valors observats pel tractament és un bon estimador de l'efecte causal. Serà un bon estimador si veiem que no té biaix i podem estimar la seva variància sota mostreig repetit. Per mostreig repetit ens referim a la selecció a l'atzar repetida d'una mostra de  $W$  d'entre la població  $\Omega$ . Igual que en Fisher, els resultats potencials els considerem fixos, de forma que l'única variable aleatòria és  $W$ , que en els experiments completament aleatoritzats té una distribució coneguda, que al seu torn indueix aleatorietat a  $Y^{\text{obs}}$ .

L'efecte causal promig de la població que volem estimar el definirem com:

$$\tau = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) = \overline{Y}(1) - \overline{Y}(0),$$

on  $\overline{Y}(1) = (\sum_{i=1}^N Y_i(1))/N$ . Igual per  $\overline{Y}(0)$ .

Com ja hem vist en l'exemple de l'aspirina, gràcies a l'aleatorietat de l'assignació, un estimador natural per  $\tau$  és  $\hat{\tau}^{dif}$ :

$$\hat{\tau}^{dif} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}.$$

El següent teorema ens diu que aquest estimador és precisament un estimador sense biaix de l'efecte causal que volem estimar:

**Teorema 3.2.** *L'estimador  $\hat{\tau}^{dif}$  és un estimador sense biaix de  $\tau$ .*

**Demostració.** Podem reescriure l'estimador com:

$$\hat{\tau}^{dif} = \frac{1}{N} \sum_{i=1}^N \left( \frac{W_i Y_i(1)}{N_t/N} - \frac{(1 - W_i) Y_i(0)}{N_c/N} \right)$$

ja que si  $W_i = 1$ , llavors  $Y_i^{\text{obs}} = Y_i(1)$  i si  $W_i = 0$ ,  $Y_i^{\text{obs}} = Y_i(0)$ .

Com que considerem  $Y(0)$  i  $Y(1)$  fixats, l'únic component aleatori de l'estadístic és  $W$ . En fer l'esperança pels diferents valors de  $W$  ens queda:

$$E(\hat{\tau}^{dif}) = \frac{1}{N} \sum_{i=1}^N \left( \frac{E(W_i) Y_i(1)}{N_t/N} - \frac{E(1 - W_i) Y_i(0)}{N_c/N} \right) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) = \tau$$

que és precisament la definició que un estadístic no tingui biaix.

□

Un cop tenim que l'estimador no és esbiaixat volem veure quina és la seva variància mostral i si podem trobar un estimador per aquesta.

Denotem per  $S_t^2$  i per  $S_c^2$  les variàncies dels vectors reals dels resultats potencials  $Y_i(0)$  i de  $Y_i(1)$  que no coneixem, és a dir:

$$S_c^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(0) - \bar{Y}(0))^2, \quad \text{i} \quad S_t^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - \bar{Y}(1))^2$$

i per  $S_{tc}^2$  la variància de l'efecte del tractament promig versus l'efecte del tractament real que no coneixem, en cada unitat, és a dir:

$$S_{tc}^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - (\bar{Y}(1) - \bar{Y}(0)))^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - \tau)^2$$

Aleshores com es podria intuir, es pot demostrar que la variància mostral de  $\hat{\tau}^{dif}$  pels diferents valors de  $W$  depèn d'aquestes tres variàncies. Ometrem la demostració degut a la seva llargada, es pot trobar a “Imbens (2015)”.

**Teorema 3.3.** *La variància mostral de  $\hat{\tau}^{dif}$  és:*

$$\mathbb{V}(\hat{\tau}^{dif}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N} \quad (3.1)$$

Però com que només hem observat uns quants resultats potencials, no coneixem ni  $Y_i(1)$  i  $Y_i(0)$ , de forma que no podem calcular el valor real de  $S_t^2$ ,  $S_c^2$  i  $S_{tc}^2$ . Com que només tenim uns quants valors de  $Y_i(1)$  i de  $Y_i(0)$ , un estimador natural d'aquestes variàncies seria calcular-les igual però utilitzant només els valors de què disposem. Aquest és un recurs que s'utilitza molt sovint en la teoria de mostreig aleatori simple. Així, proposem com estimadors de  $S_t^2$  i  $S_c^2$ , respectivament a:

$$s_t^2 = \frac{1}{N_t-1} \sum_{i:W_i=1} (Y_i^{obs} - \bar{Y}_t^{obs})^2$$

$$s_c^2 = \frac{1}{N_c-1} \sum_{i:W_i=0} (Y_i^{obs} - \bar{Y}_c^{obs})^2$$

El tercer terme de 3.1,  $S_{tc}^2$  no el podem estimar d'aquesta forma perquè mai observem ambdós valors  $Y_i(0)$  i  $Y_i(1)$  per a la mateixa unitat. De fet, no tenim cap forma d'observar directament la variació de l'efecte del tractament entre la població i per tant, no podem estimar directament  $S_{tc}^2$ . L'única opció que tenim és suposar que l'efecte del tractament és constant al llarg de la població de manera que aquest terme s'anul·larà.

Així doncs, sota aquesta suposició hem trobat una estimació sense biaix de 3.1:

**Teorema 3.4.** *Si l'efecte del tractament  $Y_i(1) - Y_i(0)$  és constant, aleshores un estimador sense biaix de la variància mostral de  $\hat{\tau}^{dif}$  és:*

$$\hat{\mathbb{V}}^{neyman} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$$

En el cas que volguéssim incloure les covariants en l'anàlisi per incrementar la precisió de l'estimació, el que podem fer és dividir la mostra en estrats segons valors iguals de les covariants i calcular l'estadístic  $\hat{\tau}^{dif}$  per a cada estrat, que sabrem pel que hem vist que és un estimador sense biaix de l'efecte causal promig  $\tau$  en el si de l'estrat. Alternativament, si optéssim per la via de l'últim capítol, canviant l'estadístic  $\hat{\tau}^{dif}$  per un que depengués també de les covariants, seria molt difícil comprovar que és un estimador sense biaix i donar una estimació també sense biaix de la seva variància. El mètode de Neyman és poc flexible respecte a canvis en l'especificació de l'efecte causal.

### 3.4 Mètodes de regressió

Una altra forma de trobar l'efecte causal en el context d'un estudi experimental completament aleatoritzat és usant mètodes que utilitzin regressions lineals. Les regressions lineals involucrades són funcions que descriuen els resultats potencials observats en termes de l'indicador del tractament. Per exemple, podem considerar la regressió lineal simple:

$$Y_i^{\text{obs}} \approx \alpha + \tau W_i \quad (3.2)$$

Si anomenem  $\epsilon_i$  al residu d'aquesta aproximació, la podem reescriure com:

$$Y_i^{\text{obs}} = \alpha + \tau W_i + \epsilon_i$$

Si poguéssim assumir que  $W_i$  fos independent de  $\epsilon_i$ , o almenys que no estiguessin correlacionats, és clar que el paràmetre  $\tau$  representaria l'efecte causal del tractament, ja que quantificaria la dependència de  $Y_i^{\text{obs}}$  respecte  $W_i$ . Generalment, és molt difícil verificar aquesta suposició directament. En el nostre context d'un experiment completament aleatoritzat, esquivarem aquesta suposició veient que l'estimador de  $\tau$  és un bon estimador per a l'efecte causal del tractament. L'estimació de  $\tau$  la trobarem utilitzant el mètode dels mínims quadrats, en el qual els estimadors de  $\alpha$  i  $\tau$ , que anomenarem  $\hat{\alpha}^{\text{mq}}$  i  $\hat{\tau}^{\text{mq}}$ , són els paràmetres que minimitzen la suma dels residus quadràtics:

$$(\hat{\tau}^{\text{mq}}, \hat{\alpha}^{\text{mq}}) = \arg \min_{\alpha, \tau} \sum_{i=1}^N (Y_i^{\text{obs}} - \alpha - \tau W_i)^2$$

La ben coneguda solució d'aquesta equació és:

$$\hat{\tau}^{\text{mq}} = \frac{\sum_{i=1}^N (W_i - \bar{W})(Y_i^{\text{obs}} - \bar{Y}^{\text{obs}})}{\sum_{i=1}^N (W_i - \bar{W})^2}, \quad \text{i} \quad \hat{\alpha}^{\text{mq}} = \bar{Y}^{\text{obs}} - \hat{\tau}^{\text{mq}} \bar{W}, \quad (3.3)$$

on:

$$\bar{W} = \frac{1}{N} \sum_{i=1}^N W_i = \frac{N_t}{N} \quad \text{i} \quad \bar{Y}^{\text{obs}} = \frac{1}{N} \sum_{i=1}^N Y_i^{\text{obs}}$$

**Proposició 3.5.** *L'estimador que és solució de l'equació de mínims quadrats coincideix amb l'estimador promig de les diferències observades:*

$$\hat{\tau}^{mq} = \overline{Y}_t^{obs} - \overline{Y}_c^{obs} = \hat{\tau}^{dif}$$

**Demostració:** Partim el sumatori del numerador de  $\hat{\tau}^{mq}$  per les unitats tractades i pels controls i fem  $\overline{W} = \frac{N_t}{N}$ :

$$\begin{aligned} \sum_{i=1}^N (W_i - \overline{W})(Y_i^{obs} - \overline{Y}^{obs}) &= \sum_{i:W_i=1} (1 - \frac{N_t}{N})(Y_i^{obs} - \overline{Y}^{obs}) + \sum_{i:W_i=0} (\frac{-N_t}{N})(Y_i^{obs} - \overline{Y}^{obs}) \\ &= \sum_{i:W_i=1} \frac{N_c}{N} (Y_i^{obs} - \overline{Y}^{obs}) - \sum_{i:W_i=0} \frac{N_t}{N} (Y_i^{obs} - \overline{Y}^{obs}) \\ &= \frac{N_c}{N} (N_t \overline{Y}_t^{obs} - N_t \overline{Y}^{obs}) - \frac{N_t}{N} (N_c \overline{Y}_c^{obs} - N_c \overline{Y}^{obs}) \\ &= \frac{N_c N_t}{N} (\overline{Y}_t^{obs} - \overline{Y}_c^{obs}) \end{aligned}$$

Partim ara el sumatori del denominador d'igual forma:

$$\begin{aligned} \sum_{i=1}^N (W_i - \overline{W})^2 &= \sum_{i:W_i=1} \left(1 - \frac{N_t}{N}\right)^2 + \sum_{i:W_i=0} \left(\frac{-N_t}{N}\right)^2 = \sum_{i:W_i=1} \left(\frac{N_c}{N}\right)^2 + \sum_{i:W_i=0} \left(\frac{N_t}{N}\right)^2 \\ &= \frac{N_t N_c^2 + N_c N_t^2}{N^2} = \frac{N_c N_t}{N} \left(\frac{N_c + N_t}{N}\right) = \frac{N_c N_t}{N} \end{aligned}$$

D'aquesta manera, substituint el numerador i el denominador de  $\hat{\tau}^{mq}$  per les seves noves expressions, ens queda de la forma:

$$\hat{\tau}^{mq} = \frac{\sum_{i=1}^N (W_i - \overline{W})(Y_i^{obs} - \overline{Y}^{obs})}{\sum_{i=1}^N (W_i - \overline{W})^2} = \overline{Y}_t^{obs} - \overline{Y}_c^{obs}$$

que és la definició de  $\hat{\tau}^{dif}$ .

□

Així doncs, com que tenim que  $\hat{\tau}^{mq}$  coincideix amb  $\hat{\tau}^{dif}$ , podem interpretar causalment  $\tau$  sense fer cap suposició extra perquè  $\hat{\tau}^{dif}$  és un estimador sense biaix de l'efecte causal en el context d'un experiment completament aleatoritzat, com hem vist en l'anterior capítol.

Si volem incrementar la precisió de l'estimació controlant per les covariants  $X$ , les podem incloure a la regressió:

$$Y_i^{obs} = \alpha + \tau W_i + X_i \beta + \epsilon_i$$

Estimem els paràmetres utilitzant el mètode dels mínims quadrats un altre cop:

$$(\hat{\tau}^{mq}, \hat{\alpha}^{mq}, \hat{\gamma}^{mq}) = \arg \min \sum_{i=1}^N (Y_i^{obs} - \alpha - \tau W_i - \gamma X_i)^2$$

Notem que l'única estimació que ens interessa és  $\hat{\tau}^{\text{mq}}$ . Obtindríem els mateixos resultats que teníem per a la regressió inicial utilitzant els mateixos arguments.

Per incrementar la precisió del model, si considerem que l'associació de les covariants amb els resultats potencials pot variar amb el tractament podem incloure a la regressió les interaccions de les covariants amb el tractament de la següent manera:

$$Y_i^{\text{obs}} = \alpha + \tau W_i + X_i \beta + \gamma W_i X_i + \epsilon_i$$

Tornaríem a obtenir, d'igual forma, els mateixos resultats vistos per a la primera regressió.

### 3.5 Inferència utilitzant models

Totes les formes que hem vist fins ara per inferir l'efecte del tractament acaben sent implícitament mètodes d'imputació dels resultats potencials que falten  $Y^{\text{fal}}$ . En aquest cas, aquest mètode és explícitament un mètode d'imputació de  $Y^{\text{fal}}$ . Es basa a trobar un model per als resultats potencials que falten, donats els resultats potencials observats:

$$f(Y^{\text{fal}} | Y^{\text{obs}}, W) \quad (3.4)$$

A partir d'aquest model deduirem la distribució de l'estimador causal d'interès, que com sempre denotarem per  $\tau$ . Podem utilitzar tota mena d'estimador de l'efecte causal, ja que, a diferència dels altres mètodes, aquest és molt flexible així que qualsevol s'adaptarà bé. En general, en l'absència de covariants l'estimador serà de la forma:

$$\tau = \tau(Y(0), Y(1), W)$$

Si considerem la següent transformació:

$$Y_i(0) = W_i Y_i^{\text{fal}} + (1 - W_i) Y_i^{\text{obs}}, \quad \text{i} \quad Y_i(1) = W_i Y_i^{\text{obs}} + (1 - W_i) Y_i^{\text{fal}}$$

podem reescriure l'estimador en la forma  $\tau = \tau(Y^{\text{fal}}, Y^{\text{obs}}, W)$ . Així doncs, conèixer 3.4 ens permetrà obtenir la distribució de l'estimador condicionat als valors observats, que és el que anomenarem distribució a posteriori. El terme distribució a posteriori és recurrent en l'estadística Bayesiana i fa referència a la rectificació, un cop hem observat els valors reals, d'una distribució hipotètica que hem plantejat inicialment que anomenem distribució a priori. Això és el que farem en els següents 4 passos per a trobar la distribució a posteriori de  $\tau$ :

**Pas 1: Trobar  $f(Y^{\text{fal}} | Y^{\text{obs}}, W, \theta)$ .** Els resultats potencials ja no ens els mirem com a fixos, sinó com a variables aleatòries. Així, començarem plantejant un model per a la probabilitat conjunta de  $Y(0), Y(1)$  condicionat a un paràmetre  $\theta$ :

$$f(Y(0), Y(1) | \theta) \quad (3.5)$$



a partir del coneixement sobre el terreny de l'experiment. Especificarem també una distribució a priori pel paràmetre  $\theta$ :

$$p(\theta)$$

Generalment, la tria d'una distribució o altra no afecta els resultats substancials derivats del mètode.

Notem que tenim la següent igualtat:

$$f(Y(0), Y(1) | \theta) = f(Y(0), Y(1) | W, \theta)$$

perquè en el context d'un experiment completament aleatoritzat  $W$  és independent de  $(Y(0), Y(1))$ .

$(Y^{\text{fal}}, Y^{\text{obs}})$  es pot escriure en termes de  $(Y(0), Y(1), W)$  de la següent manera:

$$Y_i^{\text{obs}} = W_i Y_i(1) + (1 - W_i) Y_i(0), \quad \text{i} \quad Y_i^{\text{fal}} = W_i Y_i(0) + (1 - W_i) Y_i(1)$$

de forma que podem utilitzar aquesta transformació en 3.5 per trobar la distribució conjunta de  $(Y^{\text{fal}}, Y^{\text{obs}})$  donades  $W$  i  $\theta$ :

$$f(Y^{\text{fal}}, Y^{\text{obs}} | W, \theta)$$

que al seu torn, utilitzant la definició de probabilitat condicionada reiteradament, obtenim la distribució que volíem:

$$\begin{aligned} f(Y^{\text{fal}} | Y^{\text{obs}}, W, \theta) &= \frac{f(Y^{\text{fal}}, Y^{\text{obs}}, W, \theta)}{f(Y^{\text{obs}}, W, \theta)} = \frac{f(Y^{\text{fal}}, Y^{\text{obs}} | W, \theta)}{f(Y^{\text{obs}} | W, \theta)} \\ &= \frac{f(Y^{\text{fal}}, Y^{\text{obs}} | W, \theta)}{\int_{y^{\text{fal}}} f(y^{\text{fal}}, Y^{\text{obs}} | W, \theta) dy^{\text{fal}}} \end{aligned} \quad (3.6)$$

**Pas 2: Trobar la distribució a posteriori de  $\theta$ ,  $p(\theta | Y^{\text{obs}}, W)$ .** Combinarem la distribució a priori proposada  $p(\theta)$  amb la distribució de  $Y^{\text{obs}}$  donat  $\theta$  que calcularem, per a obtenir la distribució a posteriori  $p(\theta | Y^{\text{obs}}, W)$ , que es defineix de la següent manera, seguint la fórmula de Bayes per a densitats de probabilitat:

$$p(\theta | Y^{\text{obs}}, W) = \frac{p(\theta) f(Y^{\text{obs}}, W | \theta)}{f(Y^{\text{obs}}, W)} \quad (3.7)$$

on  $f(Y^{\text{obs}}, W | \theta)$  és la funció de versemblança de  $\theta$  i es calcula com:

$$f(Y^{\text{obs}}, W | \theta) = \int_{y^{\text{fal}}} f(y^{\text{fal}}, Y^{\text{obs}}, W | \theta) dy^{\text{fal}}$$

La distribució marginal de  $(Y^{\text{obs}}, W)$ ,  $f(Y^{\text{obs}}, W)$ , es calcula al seu torn com:

$$f(Y^{\text{obs}}, W) = \int_{\theta} f(Y^{\text{obs}}, W | \theta) p(\theta) d\theta$$

**Pas 3: Trobar la distribució a posteriori dels resultats potencials que falten  $f(Y^{\text{fal}} | Y^{\text{obs}}, W)$ .** Donada la distribució a posteriori del paràmetre  $\theta$  calculada en 3.7 i la distribució condicionada de  $Y^{\text{fal}}$  donats  $(Y^{\text{obs}}, W, \theta)$  en 3.6, obtenim la distribució condicionada de  $(Y^{\text{fal}}, \theta)$  donats  $(Y^{\text{obs}}, W)$ , aplicant reiteradament la definició de la distribució condicionada:

$$f(Y^{\text{fal}}, \theta | Y^{\text{obs}}, W) = f(Y^{\text{fal}} | Y^{\text{obs}}, W, \theta) p(\theta | Y^{\text{obs}}, W)$$

Integrem aquesta distribució respecte de  $\theta$  per trobar la distribució que volíem:

$$f(Y^{\text{fal}} | Y^{\text{obs}}, W) = \int_{\theta} f(Y^{\text{fal}}, \theta | Y^{\text{obs}}, W) d\theta \quad (3.8)$$

**Pas 4: Trobar la distribució a posteriori de l'estimador  $f(\tau | Y^{\text{obs}}, W)$ .** Gràcies a la distribució de  $(Y^{\text{fal}} | Y^{\text{obs}}, W)$  calculada en 3.8 obtindrem la distribució a posteriori de  $\tau = \tau(Y^{\text{fal}}, Y^{\text{obs}}, W)$  donats els valors observats de la forma:

$$f(\tau | Y^{\text{obs}}, W)$$

Un cop trobada aquesta distribució, podem calcular la mitjana a posteriori de l'estimador causal  $\tau$  que podrem interpretar-la com l'efecte causal promig, la seva variància a posteriori i altres atributs de  $\tau$  que ens donaran informació sobre l'efecte causal del tractament.

## 4 Estudis observacionals

### 4.1 Assignació no-confosa

La majoria dels estudis amb els que ens trobem són estudis observacionals, en els quals ja no podem controlar el disseny del mecanisme d'assignació: ens vénen donats els individus que estan sota el tractament i els que estan sota el control. Així, com ja hem explicat, pot ser que en els dos grups hi hagi diferències substancials que afectin la variable sobre la qual volem estudiar l'efecte del tractament, de forma que l'efecte causal del tractament resultant de comparar els dos grups tingui un biaix. Per tant, en aquests estudis la propietat de no-confusió no és trivial, com sí que ho era en els estudis experimentals, on podem aleatoritzar l'assignació. Com ja hem introduït abans, serà controlant per covariants corresponents als atributs que diferenciïn els dos grups, com aconseguirem que el mecanisme d'assignació sigui no-confós. Ressaltem que tot i diferir els dos grups del tractament i el control segons unes covariants, podria ser que aquestes covariants no afectessin els resultats potencials de manera que aleshores vindria donada de facto la propietat de no-confusió.

Recordem que podíem escriure la propietat de no-confusió com:

$$W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid X_i, \quad \forall i \in \{1, \dots, N\} \quad (4.1)$$

És a dir, hem de veure que cada un dels resultats potencials té la mateixa distribució pels controls que per les unitats tractades. Però aquesta propietat és impossible de comprovar directament, ja que per aquells que han rebut el tractament, no tenim informació de la distribució del resultat potencial corresponent al control, i al revés, de forma que la propietat de no-confusió no és comprovable directament. Normalment però, si l'estudi observacional està ben dissenyat, haurem observat totes les covariants que puguin afectar els resultats potencials, de forma que podrem suposar que el mecanisme d'assignació és no-confós. El coneixement sobre el terreny que tinguem de l'estudi serà clau per sustentar aquesta suposició.

Suposem doncs que el mecanisme d'assignació satisfà la propietat de no-confusió, a més a més de la propietat individualística i probabilística. És a dir, tenim un mecanisme d'assignació regular. Aleshores, la idea és partir la població segons valors de covariants semblants de forma que dintre d'aquests grups tindrem un experiment completament aleatoritzat. Imaginem que existeix una subpoblació d'unitats amb un valor idèntic  $X_i = x$ . Aleshores, sota la hipòtesi que es compleix 4.1, les unitats tractades i els controls són mostres que han estat generades aleatòriament d'entre els elements d'aquesta subpoblació. Així doncs, dintre d'aquesta subpoblació ens podem mirar les dades com si vinguessin d'un experiment completament aleatoritzat.

Idealment, voldríem construir grups amb moltes unitats i amb atributs molt semblants (valors de les covariants gairebé idèntics). En la pràctica, haurem de sacrificar una cosa o l'altra: com més equilibri imposem, menys unitats tindrem i al revés. Ens sentim com Ulisses travessant l'estret pas marítim, forçats a triar entre passar per la vora del monstre de set caps Escil·la o bé, a prop del remolí gegant Caribdis.

Segons ens decantem a preservar una cosa o l'altra tenim dos mètodes diferents. En el primer, la partició segons el valor de propensió, intentarem mantenir un nombre suficient d'unitats a cada grup i dintre d'aquesta limitació, intentarem que estigui el màxim d'equilibrat possible. Per contra, en el “matching” (**emparellament**), els grups només tindran dues unitats per garantir l'equilibri màxim.

## 4.2 Estratificació pel valor de propensió

Si tenim moltes covariants, serà molt difícil classificar la població segons els valors de cada una. Necessitem una funció de menys dimensions que depengui d'aquestes i que sigui suficient controlar pels seus valors per tal de garantir la no-confusió:

**Definició 4.1.** *Un “balancing score” (valor d'equilibratge)  $b(x)$  és una funció de les covariants tal que satisfà*

$$W_i \perp\!\!\!\perp X_i \mid b(X_i) \quad (4.2)$$

Notem que tot i que donat un mateix valor del valor d'equilibratge les covariants puguin ser diferents, 4.2 ens assegura que la distribució dels valors de les covariants en els dos grups serà la mateixa, igual que teníem de forma directa en els experiments completament aleatoritzats. Així doncs, si el mecanisme d'assignació és no-confós controlant per  $X_i$ , ho serà també controlant per  $b(X_i)$ . Vegem-ho:

**Proposició 4.2.** *Suposem que el mecanisme d'assignació és no-confós. Llavors, el mecanisme d'assignació és no-confós donat qualsevol valor d'equilibratge:*

$$W_i \perp\!\!\!\perp Y_i(0), Y_i(1) \mid b(X_i)$$

**Demostració.** És equivalent a veure:

$$P(W_i = 1 \mid Y_i(0), Y_i(1), b(X_i)) = P(W_i = 1 \mid b(X_i))$$

Iterant esperances segons la propietat A.1 de l'Annex, el primer terme de la igualtat el podem escriure com:

$$\begin{aligned} P(W_i = 1 \mid Y_i(0), Y_i(1), b(X_i)) &= E[W_i \mid Y_i(0), Y_i(1), b(X_i)] \\ &= E[E[W_i \mid Y_i(0), Y_i(1), X_i, b(X_i)] \mid Y_i(0), Y_i(1), b(X_i)] \end{aligned} \quad (4.3)$$

Com que el mecanisme d'assignació és no-confós i  $b(x)$  és un valor d'equilibratge:

$$E[W_i \mid Y_i(0), Y_i(1), X_i, b(X_i)] = E[W_i \mid X_i, b(X_i)] = E[W_i \mid b(X_i)]$$

Per tant, 4.3 ens queda com:

$$E[E[W_i \mid b(X_i)] \mid Y_i(0), Y_i(1), b(X_i)] = E[W_i \mid b(X_i)] = P(W_i = 1 \mid b(X_i))$$

que és el segon terme de la igualtat al que volíem arribar.

□

Per tant, serà suficient classificar la població segons el valor d'equilibratge per tenir un experiment completament aleatoritzat en el si de cada bloc.

Notem que per definició els valors d'equilibratge no són únics, el propi  $X_i$  n'és un. Pel nostre propòsit però, estem interessats en funcions unidimensionals. Així doncs, utilitzarem com a valor d'equilibratge el valor de propensió, que recordem que quan el mecanisme d'assignació és regular, denota la probabilitat d'una unitat de ser-li assignat el tractament, donat el seu valor de les covariants:

$$e(x) = P(W_i = 1 | X_i = x) = \mathbb{E}[W_i | X_i = x]$$

Veiem que satisfà la propietat dels valors d'equilibratge:

**Proposició 4.3.** *El valor de propensió és un valor d'equilibratge*

**Demostració.** Que el valor de propensió compleixi 4.2 és equivalent a:

$$P(W_i = 1 | X_i, e(X_i)) = P(W_i = 1 | e(X_i))$$

El terme de l'esquerra ens queda:

$$P(W_i = 1 | X_i, e(X_i)) = P(W_i = 1 | X_i) = e(X_i)$$

perquè el valor de propensió és funció de les covariants. I el terme de la dreta, iterant esperances tornant a utilitzar la propietat A.1 de l'Annex:

$$P(W_i = 1 | e(X_i)) = E[W_i | e(X_i)] = E[E[W_i | X_i, e(X_i)] | e(X_i)] = E[e(X_i) | e(X_i)] = e(X_i)$$

□

Però precisament en el context d'un estudi observacional, no coneixem la probabilitat de cada unitat de ser-li assignat el tractament, és a dir, no coneixem el valor de propensió. La primera dificultat serà doncs trobar-ne una estimació, per després poder classificar la població en blocs segons el valor d'aquesta.

#### 4.2.1 Estimació del valor de propensió

Cal tenir en compte primer que el valor de propensió no té per què dependre de totes les covariants presents en l'estudi. També pot passar que per alguna de les covariants més importants, no sigui suficient incloure-les linealment sinó que hàgim de tenir en compte ordres superiors, com el seu terme quadràtic o interaccions amb alguna altra covariant. Tenint això present, especificarem un model de regressió logística de l'indicador del tractament  $W_i$  en funció de les covariants de  $X_i$  que facin falta, linealment i quadràticament, i intentarem estimar els seus coeficients per màxima versemblança. Com que el valor de propensió és precisament la probabilitat que  $W_i = 1$  en funció de les covariants, aquesta regressió serà una forma d'estimar-lo. Descrivim el procés de selecció de les covariants que participaran:

Considerem el vector de les covariants  $X_i$  de  $K$  components. D'aquestes  $K$ , triem un número  $K_B$  de covariants bàsiques que creiem que són importants a l'hora de

determinar el valor de propensió, basant-nos en informació que tenim a priori sobre l'estudi. A continuació, fem per cada una de les  $K - K_B$  covariants restants una regressió lineal afegint la respectiva covariant i veient si és prou significativa per explicar  $W_i$ . És a dir, en el cas que alguna d'aquestes regressions tingui un valor de significació més gran que una consant  $C_L$  prefixada, afegirem el que el tingui més gran a la regressió que ja té  $K_B$  covariants. Farem exactament el mateix procediment per a les  $K - (K_B + 1)$  restants, i així successivament fins que ens trobem en un pas en el qual tots els valors de significació de les covariants que puguem afegir siguin tots menors de  $C_L$ . Al final del procés, haurem afegit  $K_L - K_B$  covariants a les  $K_B$  covariants que ja havíem seleccionat, de manera que un total de  $K_L$  covariants entraran linealment en el model.

Un cop seleccionats els  $K_L$  termes lineals, hem de decidir quines interaccions i quins termes quadràtics involucrant aquestes  $K_L$  covariants afegirem al model sobre el valor de propensió. Seguirem un procediment semblant a l'anterior, considerem tots els termes quadràtics i totes les interaccions possibles de les  $K_L$  covariants, que són  $K_L(K_L + 1)/2$ . Suposem que en un punt hem afegit  $\tilde{K}_Q$  termes de segon ordre, per a cada  $K_L(K_L + 1)/2 - \tilde{K}_Q$  restant, estimem el seu coeficient en una regressió que inclou aquest terme quadràtic més els  $K_L$  termes lineals i les  $\tilde{K}_Q$  seleccionats fins aquest punt. Si el valor de significació més gran és major que una constant prefixada  $C_Q$ , inclourem el terme en el model i tornarem a fer el mateix procediment però ara amb  $\tilde{K}_Q + 1$  termes de segon ordre seleccionats.

El model de regressió obtingut amb els  $K_L + K_Q + 1$  termes corresponents (incloem el coeficient d'intercepció) serà una estimació del valor de propensió que denotarem per  $\hat{e}(x)$ , funció de cada valor de les covariants  $x$ .

#### 4.2.2 Subclassificació en estrats equilibrats

Donada l'estimació del valor de propensió  $\hat{e}(x)$ , volem partir la població en blocs (o estrats) que tinguin un valor semblant de  $\hat{e}(x)$ , de forma que puguem conduir un experiment completament aleatoritzat dins de cada bloc, sempre que tinguem un mecanisme d'assignació regular. La primera dificultat és assegurar-nos que dintre de cada bloc els valors de l'estimador estiguin equilibrats entre el grup del tractament i el del control.

Suposem que hem seleccionat un conjunt de valors que conformaran les fronteres dels blocs  $0 = b_0 < b_1 < \dots < b_{J-1} < b_J = 1$ , definim l'indicador del bloc  $B_i(j)$ , que ens determinarà si la unitat  $i$ -èsima pertany al bloc  $j \in \{1, \dots, J\}$ , com:

$$B_i(j) = \begin{cases} 1, & \text{si } b_{j-1} \leq \hat{e}(X_i) < b_j, \\ 0, & \text{altrament} \end{cases}$$

Triarem els valors  $b_j$  que delimiten els blocs de forma que per a cada estrat  $\hat{e}(x)$  variï poc entre els dos grups adequant-se al nostre propòsit. Per a veure-ho, és suficient preguntar-nos si dintre de cada bloc l'estimador és independent de  $W_i$ .

Farem servir la linearització de l'estimador del valor de propensió:

$$\hat{l}(x) = \ln\left(\frac{\hat{e}(x)}{1 - \hat{e}(x)}\right)$$

Siguin  $\bar{l}_c(j)$  i  $\bar{l}_t(j)$  els valors promigs de l'estimador del valor de propensió dels controls i de les unitats tractades respectivament, dins del bloc  $j$ -èssim. Comprovarem la independència de  $\hat{l}(x)$  respecte de  $W_i$  fent un t-test per a cada estrat sobre l'estadístic:

$$t_j = \frac{\bar{l}_t(j) - \bar{l}_c(j)}{\sqrt{S_l^2(j)(1/N_c(j) + 1/N_t(j))}} \quad (4.4)$$

on  $S_l^2(j)$  és la variància mostral del valor de propensió linearitzat en el bloc  $j$ -èssim:

$$S_l^2(j) = \frac{1}{N(j) - 2} \left( \sum_{i:B(j)=1}^N (1 - W_i) (\hat{l}(X_i) - \bar{l}_c(j))^2 + \sum_{i:B(j)=1}^N W_i (\hat{l}(X_i) - \bar{l}_t(j))^2 \right)$$

Contrastem la hipòtesi nul·la que  $\bar{l}_t(j) = \bar{l}_c(j)$ . Assumint que la distribució dels promigs  $\bar{l}_t(j)$  segueix una distribució normal, si estigués centrada en  $\bar{l}_c(j)$  com volem veure, aleshores  $\bar{l}_t(j) - \bar{l}_c(j)$  seguiria una distribució normal centrada en el zero. Com que el denominador de  $t_j$  és precisament l'error estàndard d'aquesta diferència de promitjos, sota la hipòtesi nul·la  $t_j$  seguiria una distribució  $N(0, 1)$  pel teorema central del límit. Escollirem un nivell de significació pel test:  $t_{\max}$ , a partir del qual considerarem poc probable el valor de  $t_j$  en el cas que seguís una  $N(0, 1)$ . Per tant, si  $|t_j|$  no supera aquest valor  $t_{\max}$  prefixat per a tot  $j \in \{1, \dots, J\}$  podrem concloure que l'estimador del valor de propensió varia poc dintre del bloc. Si per contra, algun bloc  $j$  supera  $t_{\max}$ , hauríem de partir-lo en dos nous blocs. Quin serà doncs el nou valor de les fronteres del bloc? Si anomenem  $b'_j$  a la mediana de l'estrat (s'agafa la mitjana dels dos valors centrals si el nombre d'unitats de l'estrat és parell) els dos nous blocs seran:  $(b_{j-1}, b'_j)$  i  $(b'_j, b_j)$ , sempre que el nombre d'unitats per a cada nou estrat sigui suficient gran. Per suficientment gran, demanarem que el grup de control i de tractament de cada bloc superi un nombre mínim d'unitats que denotarem per  $N_{\min}$ , és a dir:  $\min_j(N_c(j), N_t(j)) \geq N_{\min}$ .

Iterarem aquest procediment fins que tots els blocs tinguin unitats amb valors prou semblants de  $\hat{e}(x)$  o siguin massa petits per ser partits en dos. Podem agafar com a partició inicial d'aquest procés iteratiu un sol bloc que englobi totes les unitats.

Normalment utilitzarem  $t_{\max} = 1$  i  $N_{\min} = 10$ . Com més petit sigui  $t_{\max}$  més difícil serà que els blocs superin  $N_{\min}$ , i al revés.

Un cop tenim la població subclassificada en estrats equilibrats segons el valor de propensió, hem de veure que podem aproximar el mecanisme d'assignació per un experiment completament aleatoritzat en el si de cada estrat. En principi, com que  $\hat{e}(x)$  no és el valor de propensió real, sinó una estimació, no tindria per què ser un valor d'equilibratge i per tant verificar la propietat que ens permetria fer aquesta aproximació:

$$W_i \perp\!\!\!\perp X_i \mid \hat{e}(X_i)$$

Així doncs, ens hem d'assegurar que per cada bloc resultant tinguem que  $W_i$  és independent de  $X_i$ . Donada la partició proposada, és equivalent a veure que  $W_i$  i  $X_i$  siguin independents dins de cada estrat:

$$W_i \perp\!\!\!\perp X_i \mid B_i(1), \dots, B_i(J)$$

que és el mateix que veure que dintre de cada bloc les covariants estan equilibrades entre els dos grups. Una forma de veure aquest equilibri és la següent:

Prenem la component  $k$ -èssima del vector de covariants  $X_i$ , considerem l'estadístic corresponent a la diferència de promigs de la covariant  $k$ -èssima del grup tractat respecte del control, en el si de l'estrat  $j$ -èssim :

$$\hat{\tau}_k^X(j) = \overline{X}_{t,k}(j) - \overline{X}_{c,k}(j)$$

Podem estimar la seva variança  $\hat{V}_k^X(j)$  per:

$$\hat{V}_k^X(j) = \left( \frac{1}{N_c(j)} + \frac{1}{N_t(j)} \right) s_k^2(j)$$

on:

$$s_k^2(j) = \frac{1}{N_c(j) - 2} \left( \sum_{i:B_i(j)=1}^N (1 - W_i)(X_{ik} - \overline{X}_{c,k}(j))^2 + \sum_{i:B_i(j)=0}^N W_i(X_{ik} - \overline{X}_{t,k}(j))^2 \right)$$

Si denotem per  $\hat{\tau}_k^X$  i  $\hat{V}_k^X$  als respectius promigs ponderats pel pes de cada bloc de  $\hat{\tau}_k^X(j)$  i  $\hat{V}_k^X(j)$  podem fer un test d'hipòtesi sobre l'estadístic estandarditzat:

$$z_k^X = \frac{\hat{\tau}_k^X}{\sqrt{\hat{V}_k^X}} \quad (4.5)$$

Suposarem la hipòtesi nul·la que la diferència de promigs és zero, per tant  $z_k^X$  seguirà una distribució normal centrada en el zero. Si observem que per algun  $k$ , el valor de  $z_k^X$  és més gran del que s'esperaria si seguíssim una distribució normal centrada en el zero (definirem un nivell de significació  $z_{\max}^X$ ), conclouríem que l'estratificació plantejada no porta a un bon equilibri entre les covariants, suggerint que l'especificació del valor de propensió no és l'adequada.

### 4.2.3 Anàlisi de l'efecte causal

Suposem que tenim una estratificació equilibrada de la nostra població, feta segons els plantejaments que acabem de descriure, en el context d'un estudi observacional amb un mecanisme d'assignació regular. Per tant, dins de cada bloc, tenim aproximadament un experiment completament aleatoritzat. Diem aproximadament, ja que no hem reduït del tot el biaix entre les unitats tractades i control, ja que el valor de propensió dins de cada bloc no és constant, tot i haver buscat la variació més petita possible.



Així doncs, podem utilitzar els mètodes que havíem descrit pels experiments completament aleatoritzats, però sempre tenint en compte la presència d'aquest biaix. El mètode que triarem per estimar l'efecte del tractament reduint directament aquest biaix incorporat és el model de regressió lineal amb la inclusió de les covariants. Pel restaurant  $i$ -èssim del bloc  $j$ -èssim tindrem:

$$Y_i^{\text{obs}} = \alpha(j) + \tau(j)W_i + X_i\beta(j) + \epsilon_i$$

La seva estimació pel mètode dels mínims quadrats, que denotarem per  $\hat{\tau}^{\text{mq}}(j)$ , serà una forma d'estimar l'efecte del tractament per a cada bloc, com havíem vist que passava en els experiments completament aleatoritzats. L'estimador de l'efecte global promig del tractament serà doncs:

$$\hat{\tau}^{\text{estr}} = \sum_{j=1}^J \frac{N(j)}{N} \hat{\tau}^{\text{mq}}(j)$$

En el context dels estudis experimentals, quan les distribucions de les covariants entre les unitats tractades i els controls eren iguals, la incorporació de les covariants feien més precisa l'estimació. Ara, en l'escenari plantejat, com que ja no tenim que la correlació entre les covariants i l'indicador del tractament dins del bloc sigui nul·la, ja que la partició en blocs no elimina del tot el biaix, controlar per les covariants ens ajudarà, a més a més d'incrementar la precisió, a reduir aquest biaix que sorgeix de la diferència entre les distribucions de les covariants entre tractaments i controls.

Alternativament, podríem utilitzar l'estimador promig de les diferències dels promigs de Neyman per a cada bloc. En aquest context, ja no seria un estimador sense biaix de forma que hauríem d'ajustar per covariants més enllà del mètode. Una altra forma de reduir directament el biaix de l'estimador es basaria en l'estimació utilitzant la inferència per models per a cada bloc. Veure “Imbens (2005)” per més detalls.

És molt important per aplicar aquests mètodes que tots els blocs tinguin suficients unitats. Tenint en compte això, imposarem que el nombre mínim d'unitats sigui suficientment gran.

### 4.3 Emparellament

Una altra forma de classificar la població és emparellant cada unitat d'un grup (tractament o control) amb aquella que sigui més semblant a ella de l'altre grup, segons les covariants. Com que és molt difícil que hi hagi el mateix nombre d'unitats tractades que controls, un primer inconvenient d'aquest mètode és que ignorarem algunes unitats.

Generalment no existiran parelles amb el mateix valor exacte de les covariants, necessitarem construir una distància per a poder buscar la parella més propera. Hi ha molts tipus diferents de distàncies. Triarem una o altra, segons s'adapti millor o pitjor al nostre estudi. S'haurà de tenir en compte la naturalesa de les covariants: si són contínues, discretes o simplement variables indicadores. Els resultats que tractarem a continuació són de l'anàlisi multivariable.

### 4.3.1 Distància de Mahalanobis

Sigui  $P$  una matriu de dimensió  $K \times K$ , simètrica i definida positiva. Siguin  $X_i$  i  $X_j$  els vectors de covariants de dues unitats  $i$  i  $j$ , que recordem que tenen  $K$  components. Considerarem les següents formes quadràtiques amb matriu de pes  $P$ :

$$d_P(X_i, X_j) = \sqrt{(X_i - X_j)^T P (X_i - X_j)} \quad (4.6)$$

Recordem que una matriu  $A$ , de dimensió  $n \times n$ , és definida positiva si:

$$z^T A z \geq 0 \quad \forall z \neq 0 \in \mathbb{R}^n$$

Aquesta propietat és equivalent a què  $A$  tingui els valors propis estrictament positius.

Veiem que la forma quadràtica definida en 4.6 és en efecte una distància. Com que  $P$  és simètrica, és diagonalitzable amb valors propis reals i vectors propis ortogonals, és a dir existeixen  $\lambda_1, \dots, \lambda_K \in \mathbb{R}^+$  valors propis i  $Q$  matriu ortogonal ( $Q^T = Q^{-1}$ ) amb els vectors propis com a columnes, tals que:

$$P = Q \Lambda Q^T \quad (4.7)$$

Donat un vector qualsevol  $X_i$ , considerem la seva transformada:

$$Z_i = \Lambda^{1/2} Q X_i \quad (4.8)$$

on  $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_K^{1/2})$ .

Aleshores, la distància euclidiana de dues transformades  $Z_i$  i  $Z_j$  serà equivalent a 4.6:

$$\begin{aligned} d_P(Z_i, Z_j)^2 &= ((\Lambda^{1/2} Q)(X_i - X_j))^T ((\Lambda^{1/2} Q)(X_i - X_j)) \\ &= (X_i - X_j)^T Q^T (\Lambda^{1/2})^T \Lambda^{1/2} Q (X_i - X_j) \\ &= (X_i - X_j)^T P (X_i - X_j) \end{aligned}$$

Per tant, la forma quadràtica definida en 4.6 és una distància, ja que la distància euclidiana ho és.

La distància més usada habitualment en aquests casos és la distància de Mahalanobis, que té com a matriu de pesos la inversa de la matriu de variàncies i covariàncies de cada vector columna de la matriu de covariants  $X$ . Denotem per  $C_i = (X_{1i}, \dots, X_{Ni})$ ,  $i = \{1, \dots, K\}$  la columna  $i$ -èssima de la matriu  $X$  corresponent als valors que prenen totes les unitats respecte a la covariant  $i$ -èssima. Aleshores, la matriu de variàncies i covariàncies, que denotarem per  $S$ , és:

$$S = \begin{bmatrix} \text{Var}(C_1) & \cdots & \text{Cov}(C_1, C_d) \\ \vdots & \ddots & \vdots \\ \text{Cov}(C_d, C_1) & \cdots & \text{Var}(C_d) \end{bmatrix} \quad (4.9)$$

**Definició 4.4.** La distància de Mahalanobis entre la parella de vectors de covariants de les unitats  $i$  i  $j$  és:

$$d_{S^{-1}}(X_i, X_j) = \sqrt{(X_i - X_j)^T S^{-1} (X_i - X_j)} \quad (4.10)$$

D'aquesta manera, estem donant més pes a aquelles covariants que variïn més entre les diferents unitats.

L'element  $S_{ij}$  de la matriu serà:

$$S_{ij} = \text{Cov}(C_i, C_j) = \frac{1}{N} \sum_{k=1}^N (X_{ki} - \bar{C}_i)(X_{kj} - \bar{C}_j)$$

Construïm la matriu  $S$  pas per pas. Primer, calculem la matriu dels valors centrats de  $X$ ,  $X_C$ :

$$X_c = MX, \quad \text{on} \quad M = I_N - \frac{1}{N} \mathbf{1}_N^T \mathbf{1}_N$$

amb  $\mathbf{1}_N^T = (1, \dots, 1)$ . Així,  $S$  es pot escriure com:

$$S = \frac{1}{N} X_C^T X_C$$

Aquesta expressió ens ajuda a veure que  $S$  és semidefinida positiva i per tant  $S^{-1}$  també:

$$z^T S z = \frac{1}{N} z^T X_C^T X_C z = \frac{1}{N} (X_C z)^T (X_C z) \geq 0 \quad \forall z \in \mathbb{R}^n$$

Notem doncs que  $S$  podria no ser definida positiva estricta i a més a més, podria ser que fos singular, és a dir, que no existís la seva inversa  $S^{-1}$ . Fixem-nos però que aquestes dues propietats estan lligades:

**Proposició 4.5.**  $|S| = 0 \iff \exists z \in \mathbb{R}^N \setminus \{0\} \quad \text{tal que} \quad z^T S z = 0$

**Demostració.** La implicació cap a la dreta és trivial.

Suposem que existeix un  $z$  tal que  $z^T S z = 0$ . Utilitzant la descomposició espectral 4.7 amb  $\Lambda^{1/2}$  (existeix perquè  $S$  té els valors propis positius) tindrem  $S^{1/2}$ . Aleshores:  $z^T (S^{1/2})^T S^{1/2} z = 0$ . Ara bé, com que  $S$  és simètrica podem reescriure de la següent forma:

$$(S^{1/2} z)^T S^{1/2} z = 0 \iff S^{1/2} z = 0$$

de forma que  $Sz = 0$  i per tant  $|S| = 0$ .

□

Per tant, únicament quan  $S$  sigui singular no tindrem que  $S$  sigui definida positiva i no existirà  $S^{-1}$  de forma que no podrem definir la distància de Mahalanobis directament de la forma 4.10. En aquests casos la matriu de pesos que agafarem és la pseudoinversa de Moore-Penrose de  $S$ , que denotarem per  $S^+$ . Sigui  $r = \text{rang}(S)$ ,  $\{\lambda_1, \dots, \lambda_r\}$  els valors propis no nuls amb vectors propis  $\{q_1, \dots, q_r\}$ . Denotem per

$Q_1 = [q_1 | q_2 | \dots | q_r]$  la matriu ortogonal amb aquests  $r$  vectors propis per columnes. La matriu de pesos  $S^+$  que agafarem serà:

$$S^+ = Q_1 \begin{bmatrix} \lambda_1^{-1} & & \\ & \ddots & \\ & & \lambda_r^{-1} \end{bmatrix} Q_1^T \quad (4.11)$$

i per tant la generalització de la distància de Mahalanobis serà:

$$d_{S^+} = \sqrt{(X_i - X_j)^T S^+ (X_i - X_j)}$$

**Nota.** Per raons d'estabilitat i simplificació del mètode, encara que  $S$  no sigui singular, quan la dimensió de la matriu sigui molt gran, tractarem com nuls aquells valors propis que siguin massa petits en relació amb els altres. Per tant, en aquests casos també utilitzarem com a matriu de pes la matriu  $S^+$  definida en 4.11, utilitzant només aquells valors propis  $\{\lambda_1, \dots, \lambda_r\}$  que siguin prou grans. Per exemple, només inclourem  $\lambda_i$  si:

$$\frac{\lambda_i}{\sum_{j=1}^K \lambda_j} \geq 0.05 \quad (4.12)$$

Podríem construir altres distàncies que s'adaptin millor a les nostres dades partint de la distància de Mahalanobis. Per exemple, si hi hagués una covariant que a priori, penséssim que té més importància que les altres a l'hora de determinar els resultats potencials, podríem construir la matriu  $S^{-1}$  o  $S^+$  incrementant l'element corresponent a la covariant en qüestió donant-li d'aquesta forma més pes.

### 4.3.2 Emparellament i reemplaçament

Ordenem les unitats de la mostra segons estiguin al tractament o al control de forma que tenim  $\mathbb{I}_t := \{1, \dots, N_t\}$  corresponent als índexs de les unitats tractades i  $\mathbb{I}_c = \{N_t + 1, \dots, N_t + N_c\}$  als dels controls. Suposem, per exemple, que emparellem unitats tractades amb unitats control.

Considerem la unitat  $i$ -èssima del grup del tractament. Haurem de triar-li una parella semblant en el grup de control amb índex que denotarem per  $m_i^c$ . És a dir, donada la distància de Mahalanobis  $d_{V_M}$ , exigirem que  $m_i^c$  satisfaci:

$$m_i^c = \arg \min_{i' \in \mathbb{I}_c} d_{V_M}(X_i, X_{i'}) \quad (4.13)$$

És a dir, la solució d'aquesta equació donarà l'índex de la unitat del grup de control que sigui més propera a la unitat  $i$ -èssima del tractament segons la distància de Mahalanobis.

Com que solucionarem l'equació 4.13 per a cada unitat del tractament per separat, pot passar que la mateixa unitat del control sigui seleccionada com la parella de diferents unitats del tractament. Si no permetem que una unitat del control sigui emparellada amb més d'una unitat del tractament tindrem el que s'anomena emparellament sense reemplaçament. Si ho permetem, tindrem l'emparellament amb reemplaçament.

**Emparellament sense reemplaçament.** Suposem que duem a terme un emparellament sense reemplaçament. Com que dues unitats diferents no poden tenir la mateixa parella, haurem d'emparellar les unitats del grup més petit amb les unitats del grup més gran. Si suposem que  $N_t \leq N_c$ , emparellarem el grup de tractament amb el de control. Tenim dues formes d'evitar que més d'una unitat tractada s'emparelli amb el mateix control. La primera consisteix a calcular l'índex  $m_i^c$  simultàniament per a totes les unitats del tractament per trobar les parelles òptimes per al conjunt de la població  $\mathbb{I}_t$ . Per exemple, podríem triar els  $N_t$  índexs  $\{m_1^c, \dots, m_{N_t}^c\} \in \mathbb{I}_c$  que minimitzen la suma total de les distàncies entre les parelles:

$$\arg \min_{\{m_1^c, \dots, m_{N_t}^c\} \in \mathbb{I}_c} \sum_{i=1}^{N_t} d_{V_M}(X_i, X_{m_i^c}) \quad \text{amb la condició que} \quad m_i^c \neq m_{i'}^c \quad (4.14)$$

Aquest algoritme pot ser molt difícil d'implementar si tenim moltes unitats. En aquest escenari, es procedeix de la següent forma:

En el primer pas, la primera unitat tractada és emparellada amb el control més proper a ella solucionant directament 4.13. Denotem per  $\mathbb{I}_c - \mathbb{M}_1^c$  el conjunt de les unitats control excloent el control emparellat amb la primera unitat del tractament, és a dir,  $\mathbb{M}_1^c = \{m_1^c\}$ . Així, per a la segona unitat tractada, buscarem el control més proper a ella d'entre el nou conjunt  $\mathbb{I}_c - \mathbb{M}_1^c$ :

$$m_2^c = \arg \min_{i' \in \mathbb{I}_c - \mathbb{M}_1^c} d_{V_M}(X_2, X_{i'})$$

I així successivament fins que totes les  $N_t$  unitats tractades estiguin emparellades. En el pas  $i$ -èssim tindrem:

$$m_i^c = \operatorname{argmin}_{i' \in \mathbb{I}_c - (\cup_{i''=1}^{i-1} \mathbb{M}_{i''}^c)} d_{V_M}(X_i, X_{i'}) \quad (4.15)$$

Notem que amb aquest procediment estem ignorant l'efecte que té la selecció d'una unitat control per formar parella d'una unitat tractada per a l'emparellament de les unitats tractades que venen després. De fet, canviant l'ordre de les unitats tractades tenim emparellaments diferents. Normalment es comença per les unitats tractades que a priori creiem que puguin ser més difícils de trobar-los una parella, per exemple, aquelles que tinguin un valor de propensió més gran. Recordem que com que el valor de propensió és la probabilitat de rebre el tractament, les unitats tractades tindran generalment un valor de propensió més gran que les del control. En general, es pot veure que l'emparellament no canvia massa entre ordenaments diferents. De fet, les diferències en la precisió de l'algoritme 4.15 respecte de 4.14 són normalment petites. A més a més, aquest algoritme proposat és molt més fàcil d'implementar.

Un cop tenim les  $N_t$  parelles, plantejarem l'estimador de l'efecte causal com el promig de les diferències entre els resultats observats de les dues parelles:

$$\hat{\tau}_t = \frac{1}{N_t} \sum_{i: W_i=1} (Y_i^{\text{obs}} - Y_{m_i^c}^{\text{obs}}) \quad (4.16)$$

És important fixar-nos que com que hem emparellat en una sola direcció, les unitats tractades amb els controls,  $\hat{\tau}_t$  no serà l'estimador de l'efecte causal de tota la mostra, sinó del grup de les unitats tractades només.

**Emparellament amb reemplaçament.** Ara podem tenir més d'una unitat que s'emparelli amb la mateixa unitat. Aquest tipus d'emparellament té l'avantatge que podem aplicar directament l'algoritme en 4.13 sense preocupar-nos per l'ordre. A més a més, les parelles formades seran més semblants. Per contra, té l'inconvenient que reduïrem el nombre d'unitats del control que al final s'emparellaran, llançant unitats.

Notem que aquí ens és indiferent si emparellem unitats tractades amb unitats control, o viceversa. Com que repetim unitats emparellades, és possible emparellar un grup gran d'unitats amb un de més petit. Així doncs, podríem fer l'emparellament en les dues direccions obtenint l'efecte del tractament en les unitats tractades  $\hat{\tau}_t$  per 4.16 i l'efecte del tractament en les unitats control  $\hat{\tau}_c$  de forma anàloga. Aleshores, podem prendre com l'estimador de l'efecte del tractament de tota la mostra, l'estimador promig:

$$\hat{\tau} = \frac{N_c}{N_c + N_t} \hat{\tau}_c + \frac{N_t}{N_c + N_t} \hat{\tau}_t \quad (4.17)$$

Un problema general que ens podem trobar en qualsevol emparellament, és que hi hagi dos o més controls a la mateixa distància d'una unitat tractada  $i$ , si estem per exemple emparellant unitats tractades amb controls. En aquests casos, podríem fer el promig dels resultats de tots els controls que coincideixen en distància (sigui  $\{\mathbb{M}_i^c\}$  el conjunt dels seus índexs) per estimar el resultat potencial sota el control de la unitat tractada  $i$ -èssima:

$$\hat{Y}_i(0) = \sum_{i' \in \mathbb{M}_i^c} \frac{Y_{i'}(0)}{M_i}$$

on  $M_i = \#\{\mathbb{M}_i^c\}$ . Un altre mètode seria triar una unitat de les  $M_i$  a l'atzar per a formar parella amb la unitat  $i$ -èssima. Tot i que l'estimació de  $Y_i(0)$  seria més pobre aleshores, en situacions en què tenim poques unitats del control i fem l'emparellament sense reemplaçament és millor utilitzar aquest mètode, ja que amb el primer estaríem gastant  $M_i$  controls en una única unitat tractada de forma que ens quedarien encara menys unitats control per emparellar amb la resta.

En general, serà més fàcil trobar bones parelles quan les distribucions de les covariants en el tractament i en el control siguin semblants. Així doncs, si els valors de propensió estan concentrats a prop de 0 i a prop d'1 l'emparellament serà complicat.

Suposem que hem emparellat els tractaments amb els controls amb reemplaçament o sense. És a dir, tenim  $N_t$  parelles amb índexs  $\{i, m_i^c\}$ . Voldríem comprovar si estan ben equilibrades. Per fer-ho, haurem de mirar-nos les diferències en el valor que pren cada covariant dins de la parella. Anomenarem discrepància en l'emparellament respecte de la covariant  $k$ -èssima a aquesta diferència:

$$D_{ik} = X_{ik} - X_{m_i^c k}$$

**Nota.** En el cas que agaféssim com a parella d'una unitat tractada, més d'un control quan aquests empatin en distància, aquesta definició no ens valdrà. Suposem que tenim una unitat tractada  $i$ -èssima que té més d'un control com a parella amb índexs  $\{\mathbb{M}_i^c\}$ . Sigui  $M_i = \#\{\mathbb{M}_i^c\}$ , agafarem el promig dels valors que pren la covariant en aquestes unitats per definir la discrepància:

$$D_{ik} = X_{ik} - \frac{1}{M_i} \sum_{i' \in \{\mathbb{M}_i^c\}} X_{i'k}$$

Donat  $D_{ik}$ , podem considerar la seva mitjana normalitzada per la variància de la covariant corresponent  $k$ -èssima en les unitats tractades i en els controls per a tota la mostra. És a dir:

$$z_k^D = \frac{\overline{D}_k}{\sqrt{(s_{tk}^2 + s_{ck}^2)/2}} \quad (4.18)$$

on  $\overline{D}_k = 1/N_t \sum_{i=1}^{N_t} D_{ik}$  i:

$$s_{tk}^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left( X_{ik} - \overline{X}_k^t \right)^2 \quad \text{i} \quad s_{ck}^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left( X_{ik} - \overline{X}_k^c \right)^2$$

amb  $\overline{X}_k^t = 1/N_t \sum_{i:W_i=1} X_{ik}$ . De forma anàloga per  $\overline{X}_k^c$ . Així doncs, un criteri per veure que les parelles estan ben equilibrades serà veient que el valor de 4.18 no supera un nivell de significació màxim  $z_{max}^D$  per a cada covariant  $k$ .

#### 4.4 Variància de l'estimador de l'efecte del tractament per emparellament

Suposem que hem dut a terme algun dels emparellaments descrits, hem vist que estava equilibrat i finalment, hem calculat el valor de l'estimador corresponent de l'efecte del tractament. Aquest valor però, no és descriptiu sense una mesura de la seva significança. Com que és molt difícil estimar la variància dels estimadors proposats en aquest capítol, una forma de calcular-la directament seria utilitzar un mètode de "bootstrapping". Aquest consisteix a construir noves mostres de les dades observades, triant aleatòriament i amb reemplaçament les unitats de la mostra original. Aquest remostreig, l'haurem de fer separatament per les unitats tractades i pel control. D'aquesta manera, formarem dos nous grups d'unitats, que seran diferents dels originals perquè haurem fet el remostreig amb reemplaçament. Donats aquests dos nous grups, tornant a fer l'emparellament en qüestió tindrem un nou valor del respectiu estimador. Si repetim aquest procés molts cops, la variància del conjunt de valors que pren serà una mesura de la variància real de l'estimador. Un cop tenim una estimació de la variància, podrem construir un interval de confiança que comprendrà el valor real de l'estimador de l'efecte del tractament.

Per més detall sobre "bootstrapping" veure "Efron(1979)".

## 5 Estudi causal sobre l'efecte de la pujada del salari mínim

Il·lustrem els mètodes descrits fins ara amb un exemple numèric d'un estudi observacional. Analitzarem les dades que provenen de l'estudi fet el 1993 per David Card i Alan B. Krueger sobre l'impacte de la pujada del salari mínim en el nombre de treballadors dels restaurants de menjar ràpid.

En Abril del 1992, es va aplicar una llei a l'estat estatunidenc de Nova Jersey que pujava el salari mínim dels 4,25\$ als 5,05\$ per hora. Per a avaluar el seu impacte, Card i Krueger van recollir informació corresponent a diverses característiques de molts restaurants de menjar ràpid de Nova Jersey, abans i després de la implementació de la llei, i de Pennsilvània (estat adjacent de Nova Jersey en el qual el salari mínim va seguir sent de 4,25\$ per hora). A partir d'ara escriurem NJ i PA per denotar aquests dos estats. Alguns d'aquests atributs recollits van ser el nombre total d'empleats, el promig dels seus salaris, el tipus de cadena a la què pertany el restaurant (Burger King, KFC, Wendy's, Roy Rogers), el preu dels menús que ofereixen, les hores d'obertura, el nombre de clients, etc. Els restaurants de PA conformen el control natural dels restaurants de NJ, ja que aquest últim és un estat petit amb una economia bastant lligada als estats que té a la vora. Podríem pensar que comparant directament aquestes característiques en els restaurants de NJ entre el període anterior i posterior a l'aplicació de la llei seria una bona forma de dur a terme el nostre estudi. Però l'efecte que avaluaríem d'aquesta forma estaria esbiaixat perquè no ens dependria únicament de la implementació o no de la llei sinó també dels canvis de l'economia en el temps que ha passat entre les dues observacions. De fet, la pujada del salari mínim es va fer enmig d'una recessió econòmica. Descrivim l'estudi en la terminologia que hem fet servir al llarg del treball:

Les unitats en aquest cas són els restaurants. El tractament correspon a la pujada del salari mínim. Així doncs, el grup de tractament el formen els restaurants de NJ, mentre que el de control són els situats a PA. De tots els atributs, ens centrarem en l'impacte de la pujada del salari mínim en el nombre final de treballadors. Així doncs, els resultats potencials d'un restaurant serà el nombre final de treballadors que tindria si estigués situat a un estat o altre.

Fixem-nos que l'estudi és observacional, ja que l'investigador no pot decidir l'estat al qual pertany el restaurant, ve donat. Així, a priori els restaurants d'un estat i altre poden tenir en conjunt característiques diferents, ja que el mecanisme d'assignació no ha estat aleatoritzat. Necessitem incloure com a covariants aquells atributs que creiem que a priori puguin diferir d'un grup a l'altre i que afectin al seu torn el nombre final de treballadors, per tal de controlar-les a l'hora de tenir un mecanisme d'assignació no confós. De tots els atributs recollits en l'estudi, seguint a Rubin (reprodueix aquest mateix estudi per il·lustrar els mètodes d'emparellament), inclourem només com a covariants el nombre d'empleats inicial, el salari dels treballadors inicial i el tipus de cadena a la qual pertany.

En total, l'estudi està compost per 410 restaurants. En 45 d'ells, hi ha alguna variable important que no s'ha pogut recollir el seu valor. Com que en total només



representen el 10,98% de la mostra, argument ben conegut en teoria de “missing data” que es pot trobar a “Little (2014)”, els podem descartar quedant-nos amb  $N := 365$  restaurants. D’aquests, tenim  $N_t = 293$  restaurants de NJ i només  $N_c = 72$  de PA.

Mantenim la notació que hem seguit al llarg del treball.  $Y$  denotarà el nombre final de treballadors;  $Y(0)$  i  $Y(1)$  els vectors dels resultats potencials per a tots els restaurants;  $W$  la variable indicadora del tractament tal que té per a components  $W_i = 1$  si el restaurant  $i$ -èssim ( $i = \{1, \dots, N\}$ ) pertany a NJ i  $W_i = 0$  si pertany a PA. Siguin  $C_j$  ( $j=\{1, \dots, 3\}$ ) les tres covariants:  $C_1$  correspondrà a la primera covariant nombre d’empleats inicial,  $C_2$  al salari dels treballadors inicial i  $C_3$  al tipus de cadena a la qual pertany (1 si Burger King, 2 si Kentucky Fried Chicken, 3 si Roy Rogers i 4 si Wendys). La taula següent recull el valor que prenen els primers restaurants respecte a les variables involucrades en l’estudi:

$i$	$Y$	$W$	$C_1$	$C_2$	$C_3$
1	20.00	0	34.00	5.00	4
2	35.50	0	24.00	5.50	4
3	29.00	0	70.50	5.00	1
4	36.50	0	23.50	5.00	1
5	11.00	0	11.00	5.25	2
6	8.50	0	9.00	5.00	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Fixem-nos que  $C_3$  és una variable qualitativa, per tant, per convertir-la en quantitativa haurem de definir tres noves variables indicadores  $BK$ ,  $KFC$  i  $RR$  pels tres primers tipus de cadena. És a dir,  $BK_i = 1$  si  $C_3(i) = 1$  i  $BK_i = 0$  si  $C_3(i) \neq 1$ . Igual amb els altres dos. No cal definir una quarta variable indicadora pel tipus de cadena que queda, Wendys, ja que ve determinada pels tres primers: si no és de cap d’ells és que serà de Wendys, i al revés. Així, definirem la matriu  $X$  de les covariants la que té per columnes el nombre inicial de treballadors, el salari promig inicial i les variables indicadores  $BK$ ,  $KFC$ ,  $RR$ .

Ens podem fer una primera idea vaga sobre l’impacte que volem estudiar, fent la diferència entre el nombre de treballadors finals i inicials i comparant la distribució d’aquests valors entre els dos grups. No s’observa una diferència gaire gran entre aquestes dues distribucions:

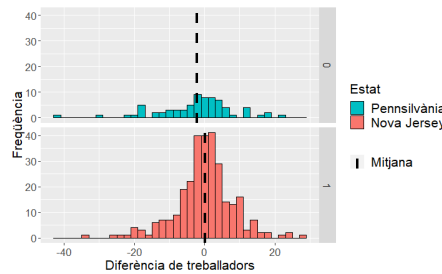


Figura 1: Comparació entre els dos grups, dels valors que pren la diferència de treballadors.

Ressaltem com es veu que a NJ tenim molts més restaurants que a PA.

Si denotem per  $\Delta E_i$  el valor de la diferència de treballadors en el restaurant  $i$ -èssim, podríem plantejar directament un mètode de regressió lineal de  $\Delta E_i$  respecte a l'indicador del tractament i les altres dues covariants:

$$\Delta E_i \approx \alpha + \beta W_i + \gamma_1 X_{i2} + \gamma_{21} BK_i + \gamma_{22} KFC_i + \gamma_{23} RR_i$$

Estimem el paràmetre  $\beta$  pel mètode dels mínims quadrats:

$\hat{\beta}^{\text{mq}}$	z-valor
2.29	1.99

Card i Krueger interpreten aquest resultat directament com l'efecte de la pujada del salari mínim, de forma que tindríem que l'impacte és positiu i a més a més, força significatiu. Però com hem vist en el treball, el paràmetre  $\beta$  del model de la regressió lineal només estima l'efecte del tractament en el context d'un experiment completament aleatoritzat, que no és el que tenim. És a dir, incloent-hi les covariants en la regressió no és suficient per eliminar la confusió que aquestes introdueixen en l'efecte del tractament per la seva condició d'estudi observacional, resultant aquesta estimació esbiaixada. Utilitzem les eines que hem descrit en el treball per tal d'aproximar l'estudi observacional en un experiment completament aleatoritzat en el que sí que podrem inferir efectes causals vàlids.

## 5.1 Assignació no-confosa

Hem fet la suposició que controlant per les covariants  $X$  definides, el mecanisme d'assignació és no-confós. És a dir, condicionant a elles, el fet que pertanyi un restaurant a un estat o a l'altre és independent del nombre final de treballadors que acabaria tenint:

$$W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid X_i, \quad \forall i \in \{1, \dots, N\}$$

En altres paraules, les covariants que tenim en compte determinen totes les diferències entre els dos grups -NJ i PA- que puguin acabar afectant el nombre final de treballadors. Il·lustrem per exemple la diferència en la distribució del nombre inicial de treballadors entre els restaurants de PA i els de NJ:

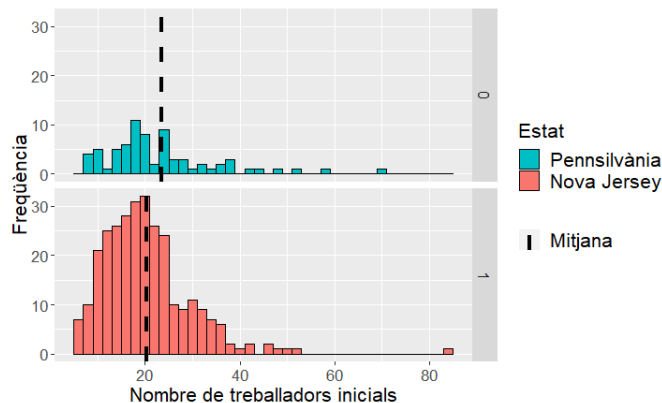


Figura 2: Distribució del nombre inicial de treballadors en els restaurants de PA i en els de NJ.

La comparació de les distribucions de les altres dues covariants entre els dos grups estan il·lustrades en la Figura 6 de l'annex.

A simple vista, notem que la diferència en les distribucions no semblen molt grans, però podrien ser suficients per introduir un biaix. Per a eliminar-lo, cal comparar restaurants de NJ amb restaurants de PA que siguin semblants. És a dir, si classifiquem per valors semblants de les covariants, reduïrem el biaix de l'estimació de l'impacte de la pujada del salari mínim (pertànyer a un estat o l'altre) en el nombre final de treballadors. Com hem vist abans, podem fer de dues formes aquesta classificació, estratificant la població en grups suficientment grans (estratificació segons el valor de propensió) o bé emparellant restaurants molt semblants (emparellament).

## 5.2 Estratificació segons el valor de propensió

### 5.2.1 Estimació del valor de propensió

Classifiquem la població en grups prou grans que tinguin un valor semblant de les covariants. No podem dur a terme aquesta estratificació directament perquè tenim 3 covariants diferents que cal controlar. Per aquesta raó, cal estimar el valor de propensió per tal de reduir-ne la dimensió i poder dur a terme aquesta estratificació.

Seguint l'algoritme proposat, plantegem per les dues primeres covariants, la regressió lineal de la variable indicadora de l'estat del restaurant  $i$ -èssim en funció del valor que pren aquesta covariant:

$$W_i \approx \alpha_k + \beta_k X_{ik} \quad k = 1, \dots, 2.$$

Pel tipus de cadena, farem la regressió incloent les tres variables indicadores:

$$W_i \approx \alpha_3 + \beta_{31} BK_i + \beta_{32} KFC_i + \beta_{33} RR_i$$

Recollim els paràmetres estimats per màxima versemblança en la taula següent, primer per la regressió de les dues primeres covariants:

$k$	$\hat{\alpha}_k$	$\hat{\beta}_k$	z-valor
1	-0.03	2.02	-2.29
2	-0.18	2.23	-0.47

j	$\hat{\beta}_{3j}$	z-val
1	0.31	0.83
2	0.90	1.91
3	0.48	1.14

I per la segona regressió:

L'única covariant significativa és el nombre d'empleats inicial, per tant serà la única que entrarà linealment en la regressió.

De l'única covariant seleccionada linealment, afegim el seu terme quadràtic a la regressió:

$$W_i \approx \alpha_1 + \beta_1 X_{i1} + \gamma_1 X_{i1}^2$$

El z-valor de  $\hat{\gamma}_1$  és 0.13 i per tant, no és significatiu.

Així doncs, l'única covariant que entra en l'estimació del valor de propensió és el terme lineal corresponent al nombre inicial de treballadors. Donat un valor  $x = (x_1, x_2, x_3)$  de les tres covariants estimarem el valor de propensió com:

$$\hat{e}(x) = \hat{e}(x_1) = -0.03 + 2.02x_1$$

D'aquesta manera, podem calcular el valor de propensió de cada restaurant en funció del nombre inicial de treballadors que té. En aquest cas, el valor de propensió no ens redueix la dimensió. Per tant, en comptes de classificar pel valor de propensió ho podríem fer directament pel nombre inicial de treballadors. Tot i això, utilitzarem el valor de propensió per il·lustrar com es faria en la gran majoria dels casos en què no només entra un únic terme en l'estimació.

### 5.2.2 Subclassificació en estrats equilibrats

Classifiquem la mostra de restaurants en grups suficientment grans dintre dels quals els restaurants dels dos grups tinguin un valor de propensió semblant. Com a suficientment grans, exigirem que cada grup tractat i control dins del bloc tingui com a mínim  $N_{min} = 10$  restaurants. Per valor semblant, demanarem que l'estadístic  $t_j^{(p)}$  definit en 4.4 no superi en valor absolut a  $t_{max} = 1$ , on  $p$  indicarà el pas del procés.

Seguint aquest criteri, comencem dividint la mostra en dos (per la mediana) perquè l'estadístic corresponent a la mostra sencera és  $t_1^{(0)} = -2.27$ . Dels nous grups formats, el que té menys unitats és el grup de control del segon bloc que és de mida 32, i per tant més gran que  $N_{min}$ , podem fer aquesta partició. Pel primer nou bloc resultant, tenim  $t_1^{(1)} = 2.09$  i pel segon,  $t_2^{(1)} = 1.19$ . Els dos blocs superen el valor màxim  $t_{max} = 1$ , per tant, tornem a dividir els dos. Comencem partint el primer bloc, el

grup més petit dels nous formats té 16 unitats. Pel primer nou bloc generat tenim  $t_1^{(2)} = 1.55$  i pel segon  $t_2^{(2)} = 0.43$ . Així doncs, aquest últim el deixem intacte, mentre que el primer l'haurem de tornar a partir. Partim ara el segon bloc, el grup més petit és de 13 unitats. El t-estadístic del primer bloc generat és  $t_3^{(2)} = 0.94$  i el del segon és  $t_4^{(2)} = -0.16$ . Per tant, no els dividirem més. En el tercer pas, partim l'únic bloc dels que hem generat en el segon pas que no ha sortit equilibrat, el primer. El grup més petit de la partició resultant té 10 unitats, igual que el mínim establert. El nou primer bloc format té  $t_1^{(3)} = 1.61$ , el segon té  $t_2^{(3)} = -0.44$ . El primer no està equilibrat. Si el partíssim ens generaria un grup que només tindria 3 unitats de forma que ens haurem de quedar amb aquest grup sencer, encara que no estigui del tot equilibrat.

D'aquesta manera, hem estratificat la mostra en 5 blocs, dels quals 4 estan molt ben equilibrats segons el valor de propensió. Ara hem de comprovar que al seu torn, estiguin ben equilibrats per a totes les covariants, per tal de poder aproximar el mecanisme d'assignació per un experiment completament aleatoritzat dins de cada bloc. Si considerem  $B_i(j)$  l'indicador que ens diu si el restaurant  $i$ -èssim pertany al bloc  $j \in \{1, \dots, 5\}$ , volem veure:

$$W_i \perp\!\!\!\perp X_i \mid B_i(1), \dots, B_i(5)$$

Per comprovar que se satisfà aquesta propietat, per cada una de les tres covariants  $k = 1, \dots, 3$ , estudiarem el valor de l'estadístic  $z_k^X$  definit en 4.5. Pren els següents valors:

$k$	$z_k^X$
1	-0.33
2	-0.23
3	-0.04
4	0.91
5	0.32

Així doncs, l'estratificació plantejada ens genera un bon equilibri entre les covariants. D'aquesta forma, podem aproximar el mecanisme d'assignació de cada estrat per un d'un experiment completament aleatoritzat.

Notem que per aquest cas, ja sabíem que cada bloc estava ben equilibrat segons la primera covariant  $C_1$ , ja que ho estava pel valor de propensió, el qual depèn únicament de  $C_1$ .

### 5.2.3 Anàlisi causal de l'impacte de la pujada del salari mínim

Hem aproximat l'estudi observacional per un experiment completament aleatoritzat per blocs. D'aquesta manera, podem aplicar per cada bloc un dels mètodes descrits en el context d'aquests experiments, corregint el biaix que no hem pogut eliminar del tot, ja que l'equilibri aconseguit per l'estratificació no és exacte. Per exemple,

considerem el següent model de regressió lineal en el si del bloc  $j$ -èssim, afegint els termes corresponents a les covariants a forma de reduir aquest biaix:

$$Y_i^{\text{obs}} \approx \alpha(j) + \tau(j)W_i + X_i\beta(j)$$

D'aquesta forma el paràmetre  $\tau(j)$  és una bona aproximació de l'efecte del tractament pels restaurants del bloc  $j$ -èssim. Utilitzant el mètode dels mínims quadrats en trobem una estimació  $\hat{\tau}^{\text{mq}}(j)$ . En la següent taula recollim el paràmetre estimat per cada bloc i el seu valor de significació:

j	$\hat{\tau}^{\text{mq}}(j)$	z-valor
1	5.87	2.15
2	1.74	0.74
3	1.37	0.72
4	-1.41	-0.79
5	-0.55	-0.32

Així, l'únic efecte del tractament significatiu és respecte a les unitats del primer bloc, que és positiu. Per la resta d'unitats l'efecte no és significatiu. Podem concloure doncs que la pujada del salari mínim afecta positivament al nombre final de treballadors en aquells restaurants amb menys empleats, que són els que tenen un valor de propensió més baix. Per a la resta de restaurants, la pujada del salari mínim no afecta el nombre final de treballadors. Una explicació hipotètica del perquè d'aquests resultats, seria que la pujada del salari mínim fes que el consum de la població augmentés, ja que la gent té més diners per gastar-se, de forma que els restaurants amb pocs treballadors haurien d'incrementar-ne el nombre, al veure augmentada la demanda.

Hem aconseguit desentrellar l'efecte del tractament que en un principi, en la regressió que utilitzaven Card i Krueger, estava esbiaixat per totes les covariants que el confonien. Però no només això, sinó que utilitzant el mètode del valor de propensió, obtenim resultats específics que abans estaven emmascarats, com per exemple que en els restaurants amb menys empleats inicials l'impacte és positiu i significatiu, a diferència dels altres. Això rau pel fet que la pujada del salari mínim no afecta a tots els restaurants per igual. Així doncs, quan ens mirem aquest efecte del tractament per a tota la mostra general, estem barrejant els diferents resultats particulars existents. Per això, pensem que l'estratificació és el millor mètode per obtenir resultats específics per cada grup diferent d'unitats.

## 5.3 Emparellament

### 5.3.1 Distància de Mahalanobis

Recordem que  $X$  denota la matriu de les covariants que comprèn el nombre total d'empleats inicials, el salari per hora inicial i les tres variables indicadores  $BK$ ,  $KFC$  i  $RR$ . Calculem la matriu de variàncies i covariàncies  $S$  d'aquesta matriu

definida en 4.9. Fem la descomposició espectral  $S = Q\Lambda Q^T$ . La matriu diagonal  $\Lambda$  és:

$$\Lambda = \begin{bmatrix} 98.22 & 0 & 0 & 0 & 0 \\ 0 & 0.33 & 0 & 0 & 0 \\ 0 & 0 & 0.19 & 0 & 0 \\ 0 & 0 & 0 & 0.11 & 0 \\ 0 & 0 & 0 & 0 & 0.04 \end{bmatrix}$$

Invertint-la directament podrem calcular  $S^{-1}$  que utilitzarem com a matriu de pes de 4.6 per calcular la distància de Mahalanobis donades dues parelles de restaurants.

### 5.3.2 Emparellament amb reemplaçament

Un cop tenim les distàncies de Mahalanobis per a cada parella de restaurants, farem l'emparellament amb reemplaçament, és a dir permetent que dues unitats s'emparellin amb una mateixa unitat. L'algoritme és molt més directe d'aquesta forma i ens permet emparellar en les dues direccions de manera que podrem trobar l'efecte del tractament promig per tota la mostra en general  $\hat{\tau}$  definit en 4.17.

Comencem calculant l'efecte del tractament per les unitats tractades  $\hat{\tau}_i$ . Per això, emparellarem els restaurants de NJ amb els de PA formant 293 parelles, els índexs de les quals els denotarem per  $\{(i, m_i^c)\}$ ,  $i = 1, \dots, 293$ . Com que hi ha molts més restaurants a NJ que a PA, per força hi haurà molts restaurants de PA que seran parella de més d'un restaurant de NJ. Si ens trobem amb més d'un restaurant de PA a la mateixa distància d'un de NJ, els tindrem en compte a tots estimant el seu resultat potencial sota el control per:

$$\hat{Y}_i(0) = \sum_{i' \in \mathbb{M}_i^c} \frac{Y_{i'}(0)}{M_i}$$

on  $\mathbb{M}_i^c$  són els índexs de tots els restaurants de PA que coincideixen en distància amb el restaurant  $i$ -èssim de NJ, i  $M_i = \#\{\mathbb{M}_i^c\}$ .

Un cop generades les parelles utilitzant aquest plantejament, calculem les discrepàncies en l'emparellament respecte a les dues primeres covariants per a totes les parelles:

$$D_{ik} = X_{ik} - X_{m_i^c k} \quad \forall i = 1, \dots, 293 \quad \text{i} \quad k = 1, 2$$

Per a les variables indicadores, farem el valor absolut d'aquesta diferència, que ens dóna 0 quan són de la mateixa cadena i 1 quan no ho són:

$$D_{ik} = |X_{ik} - X_{m_i^c k}| \quad \forall i = 1, \dots, 293 \quad \text{i} \quad k = 3, 4, 5$$

Per exemple, les discrepàncies en l'emparellament respecte al nombre inicial de treballadors es distribueixen de la següent manera entre les 293 parelles:

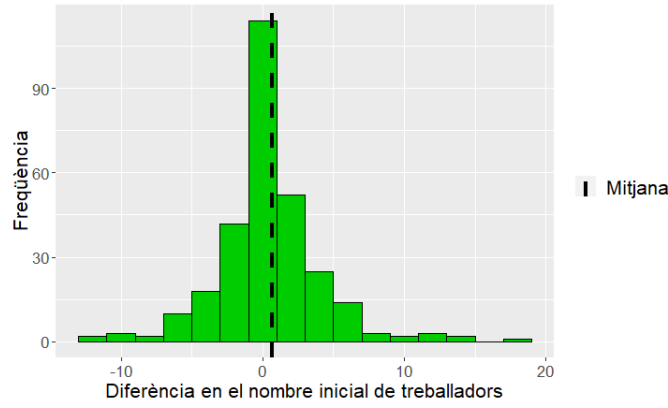


Figura 3: Discrepància en l'emparellament dels restaurants de NJ amb els de PA respecte al nombre inicial de treballadors

Les discrepàncies es distribueixen més o menys entorn del zero per tant sembla que les parelles estan ben equilibrades respecte a aquesta primera covariant. Els gràfic de les discrepàncies respecte el salari inicial és a la Figura 7 de l'Annex, i també sembla estar distribuïda entorn del zero. No hem fet l'histograma de la distribució de les discrepàncies per les variables indicadores, perquè no és gaire informatiu, ja que o bé prenen el valor 0, o bé l'1.

Calculem el valor de la mitjana normalitzada d'aquestes discrepàncies  $z_k^D$  definida en 4.18, per tal de poder valorar l'equilibri. Pren els següents valors:

k	$z_k^D$
1	0.06
2	0.02
3	0.01
4	0.00
5	0.01

Per tant, les parelles generades estan ben equilibrades respecte a totes les covariants, ja que els z-valors no surten significatius.

L'efecte del tractament pels restaurants de NJ és:

$$\hat{\tau}_t = \frac{1}{N_t} \sum_{i:W_i=1} (Y_i^{\text{obs}} - Y_{m_i^t}^{\text{obs}}) = 1.71$$

Calculem ara l'efecte del tractament pels restaurants de PA. Hem d'emparellar les unitats en l'altra direcció, és a dir, al restaurant  $i$ -èssim de PA li busquem la parella més propera d'entre els restaurants de NJ, l'índex de la qual el denotarem per  $m_i^t$ . Formarem d'aquesta manera 72 parelles  $\{i, m_i^t\}$ , on  $i = 1, \dots, 72$  són els índexs dels restaurants de PA. L'emparellament el fem de la mateixa forma que abans: amb reemplaçament i procedint de la mateixa manera quan ens trobem amb més d'una unitat a la mateixa distància.



Generades les parelles d'aquesta forma, considerem la discrepància en la parella  $i$ -èsima respecte a la covariant  $k$ -èsima, que ara vindrà donada per:

$$D_{ik} = X_{m_i^t k} - X_{ik} \quad \forall i = 1, \dots, 293 \quad \text{i} \quad k = 1, 2$$

i:

$$D_{ik} = |X_{m_i^t k} - X_{ik}| \quad \forall i = 1, \dots, 293 \quad \text{i} \quad k = 3, 4, 5$$

Per exemple, la discrepància respecte a la primera covariant, es distribueix de la següent manera entre totes les parelles:

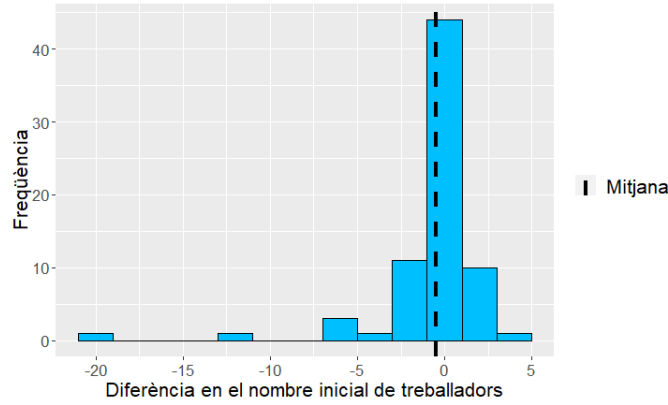


Figura 4: Discrepància en l'emparellament dels restaurants de PA amb els de NJ respecte al nombre inicial de treballadors

Estan distribuïdes al voltant del zero, igual que la discrepància respecte el salari inicial recollida en la Figura 8 de l'Annex.

Calculem els valors de  $z_k^D$  definit en 4.18 de forma idèntica a com hem fet abans i obtenim:

k	$z_k^D$
1	-0.05
2	-0.02
3	0.00
4	0.00
5	0.00

Els z-valors no són significatius, per tant estan ben equilibrades.

Calculem l'efecte del tractament sobre els restaurants de PA:

$$\hat{\tau}_c = \frac{1}{N_c} \sum_{i: W_i=0} (Y_{m_i^t}^{\text{obs}} - Y_i^{\text{obs}}) = 1.80$$

Així doncs, el promig de l'efecte del tractament sobre tota la mostra és:

$$\hat{\tau} = \frac{N_c}{N_c + N_t} \hat{\tau}_c + \frac{N_t}{N_c + N_t} \hat{\tau}_t = 1.73 \quad (5.1)$$

Mesurem la significació d'aquest valor fent “bootstrapping”. Formem 500 noves mostres cada una generades remostrejant amb reemplaçament les unitats del grup de tractament i del control, per separat. És a dir, dels 72 controls, en triem 72 a l'atzar amb reemplaçament, i de les 293 unitats fem el mateix de forma anàloga. Per cada nova mostra generada d'aquesta forma, calculem la nova distància de Mahalanobis, fem l'emparellament amb reemplaçament i calculem l'estimador de l'efecte del tractament promig  $\hat{\tau}$  com en 5.1, de la mateixa forma que hem fet per a la mostra original. Els valors de  $\hat{\tau}$  per a cada una de les 500 mostres es distribueixen de la següent forma:

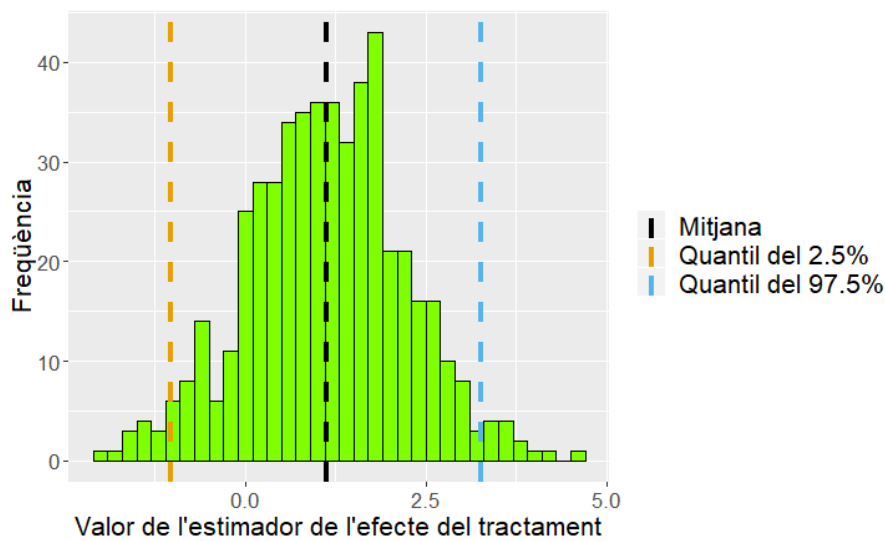


Figura 5: Distribució dels valors de  $\hat{\tau}$  per a totes les mostres generades per “bootstrapping”.

Veiem com l'interval de confiança del 95% delimitat pels quantils representats inclou el zero, per tant, podem concloure que l'efecte observat no és significatiu. Així doncs, no s'observa un efecte significatiu de la pujada del salari mínim sobre el nombre final de treballadors del global dels restaurants.

L'estimació de l'efecte a través de l'emparellament, igual que teníem en l'estratificació, és molt més acurada que la regressió lineal que duen a terme Card i Krueger perquè redueix la confusió introduïda per les covariants. Però com passa en aquesta regressió, dona una estimació promig per tota la mostra que amaga els resultats particulars no menys interessants, com el que hem trobat en l'estimació via l'estratificació. Així, pensem que si el que ens interessa és l'impacte de la pujada del salari mínim pel global de tots els restaurants de la mostra, triarem la via de l'emparellament que és molt més recte i simple. Ara, si volem conèixer l'efecte amb totes les seves variacions que pugui tenir entre el tipus de restaurant del qual parlem, triarem el camí de l'estratificació.

## 6 Conclusions

Hem vist que les eines més bàsiques de l'estadística es poden fer servir per inferir l'efecte del tractament en estudis experimentals en els quals hem aleatoritzat el mecanisme d'assignació. En la majoria dels estudis però, per la seva naturalesa observacional, aquestes eines deixen de ser vàlides i cal definir un nou marc d'actuació específic per a la inferència causal. Nosaltres, ens hem centrat únicament en el marc que principalment planteja Rubin però existeixen molts altres plantejaments i les aportacions en aquest camp estan en continu creixement.

La manera d'enfocar el problema per Rubin que hem descrit, es basa a explotar la informació sobre el terreny que tenim de l'estudi observacional per tal d'aproximar-lo per un d'experimental. Dels dos mètodes descrits, l'estratificació pel valor de propensió té l'avantatge de donar-nos l'efecte del tractament per sectors específics de la població. Pensar l'efecte del tractament de forma general és caure en l'error de suposar que la població és homogènia, quan molts cops tenim precisament el contrari. Així, tenen molt més d'interès els diferents efectes del tractament específics, que prenen en consideració les diverses particularitats dels individus entre la població.

Hem fet servir molts resultats diferents de diverses branques de l'estadística. Des de teoria de mostreig, passant per teoria de la regressió, fins a l'anàlisi multivariant.

De la diversitat de camps en els quals s'aplica la teoria de la inferència causal, hem decidit il·lustrar els nostres resultats fent un estudi sobre l'efecte de la pujada del salari mínim. D'aquesta manera, aquest treball també aporta, d'alguna forma, material al debat que té lloc en l'actualitat sobre aquesta qüestió.

## A Annex

### A.1 Propietats addicionals

Demostrem la propietat de l'esperança condicionada que hem utilitzat per a les dues demostracions corresponents als valors d'equilibratge.

**Proposició A.1.** *Siguin  $X, Y, Z$  variables aleatòries. Aleshores, se satisfà:*

$$E[Y|X] = E[E[Y|X, Z]|X]$$

**Demostració.** Suposem que les variables són discretes per tal de simplificar la demostració. Sota unes suposicions extremes, si ho tenim per les discretes, ho tindrem per les contínues.

Aleshores, per definició, la part de l'esquerra és:

$$E[Y|X] = \sum_y yP(Y = y|X = x) = \sum_y y \frac{P(X = x, Y = y)}{P(X = x)}$$

$E[Y|X, Z]$  depèn únicament del valor aleatori de  $X$  i de  $Z$  i com que estem condicionant per  $X$ , només serà funció aleatòria de  $Z$ . Per tant, la part de la dreta de la igualtat queda:

$$\begin{aligned} E[E[Y|X, Z]|X] &= \sum_w \left( \sum_y yP(Y = y|X = x, Z = z) \right) P(E[Y|X, Z] = w|X = x) \\ &= \sum_z \left( \sum_y yP(Y = y|X = x, Z = z) \right) P(Z = z|X = x) \\ &= \sum_z \left( \sum_y y \frac{P(Y = y, X = x, Z = z)}{P(X = x, Z = z)} \right) \frac{P(Z = z, X = x)}{P(X = x)} \\ &= \sum_z \left( \sum_y y \frac{P(Y = y, X = x, Z = z)}{P(X = x)} \right) \\ &= \sum_y \left( \sum_z y \frac{P(Y = y, X = x, Z = z)}{P(X = x)} \right) \\ &= \sum_y y \frac{P(X = x, Y = y)}{P(X = x)} \end{aligned}$$

que coincideix amb el que teníem per la part de l'esquerra. □

## A.2 Gràfics

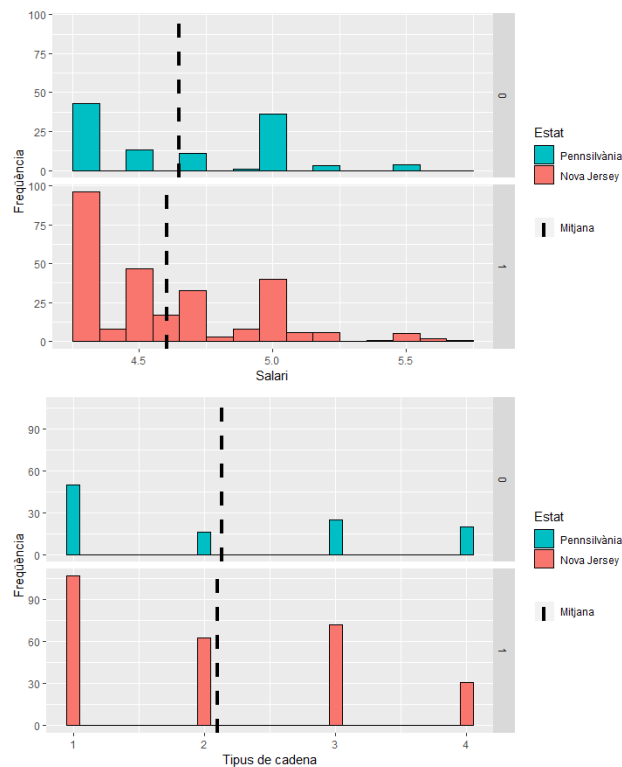


Figura 6: Diferència en els valors que prenen les covariants corresponents al salari inicial i el tipus de cadena entre els restaurants de PA i els de NJ.

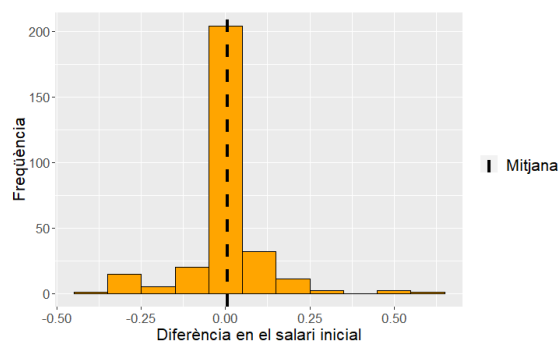


Figura 7: Discrepàncies en l'emparellament dels restaurants de NJ amb els de PA respecte al salari inicial

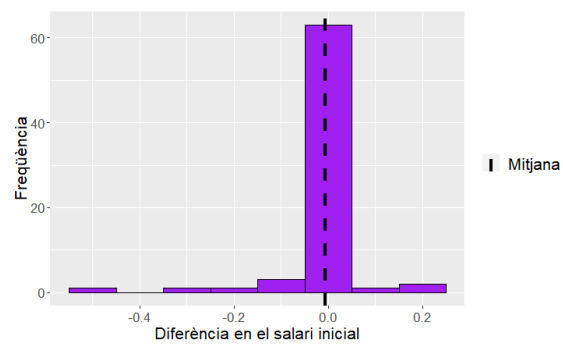


Figura 8: Discrepàncies en l'emparellament dels restaurants de PA amb els de NJ respecte al salari inicial

## Referències

- [1] Card, D., Krueger, A.B.: *Minimum Wage and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania*, National Bureau of Economic Research, Working Paper No. 4509, 1993.
- [2] Cochran, W. G.: *Sampling Techniques*, John Willey & Sons, 3a edició 1977.
- [3] Dawid, A.P.: *Conditional Independence in Statistical Theory*, Journal of the Royal Statistical Society, Series B, Vol.41(1): 1-31, 1979.
- [4] Doll, R., Hill, A.B.: *Smoking and Carcinoma of the Lung*, British Medical Journal, Vol.2(4682): 739–748, 1950.
- [5] Efron, B.: *Bootstrap Methods: Another Look at the Jackknife*, Annals of Statistics, Vol.7(1): 1-26, 1979.
- [6] Fisher, R.A.: *Smoking. The Cancer Controversy: Some Attempts to Assess the Evidence*, Oliver and Boyd, 1959.
- [7] Fisher, R.A.: *Statistical Methods for Research Workers*, Oliver and Boyd, 1925.
- [8] Imbens, G.W., Rubin, D.B.: *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press, 2015.
- [9] Little, R.J.A., Rubin, D.B.: *Statistical Analysis with Missing Data*, John Wiley & Sons, 2a edició, 2014.
- [10] Neyman, J.: *On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9*, Statistical Science, 1923, 1990.
- [11] Rubin, D.B.: *Causal Inference Using Potential Outcomes: Design, Modeling, Decisions*, Journal of the American Statistical Association, Vol.100(469):322-331, 2005.
- [12] Stolley, P.D.: *When Genius Errs: R.A. Fisher and the Lung Cancer Controversy*, American Journal of Epidemiology, Vol.133(5):416–425, 1991.